

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: October 26, 2004, 09:04:10 ; Search time 71 Seconds
(without alignments)

2536.368 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMGSLRIGATGMKTHSTGLG.....NSKSVTTADTLMQALELKR 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_23Sep04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2484	96.1	486	4	AAB36552
2	538	20.8	120	2	AAW16680
3	529.5	20.5	414	6	ABU20217
4	528.5	20.4	471	7	ABO70780
5	527.5	20.4	462	6	ABJ18752
6	506.5	19.6	413	6	ABU22839
7	491.5	19.0	402	8	ADK13784
8	484.5	18.7	411	5	ABB40983
9	483	18.7	718	2	AAW80991
10	475.5	18.4	412	7	ADF07897
11	448.5	17.3	405	6	ABM68268
12	376.5	14.6	565	4	ABG24709
13	307	11.9	269	7	ABO70369
14	296	11.4	261	6	ABM68266
15	288	11.1	271	2	AAW20373
16	288	11.1	271	2	AAW24645
17	288	11.1	279	7	AAW20863
18	286	11.1	265	7	ADF07825
19	237.5	9.2	263	3	AAW51377
20	223.5	8.6	263	2	AAW73073
21	176	6.8	2736	7	ASO81481
22	173.5	6.7	1095	3	AAW01847
23	173.5	6.7	1536	2	AAW41723
24	173.5	6.7	1536	2	AAW41725
25	173.5	6.7	1536	2	AAW63505

26	173.5	6.7	1536	3	AAW01846
27	172.5	6.7	3705	6	ABU50020
28	167.5	6.5	1536	2	AAW30293
29	166	6.4	2468	6	ABU38411
30	166	6.4	2468	6	ABP59933
31	165.5	6.4	124	2	AAW20361
32	165	6.4	5291	7	ADC01014
33	162	6.3	1954	8	ADF83256
34	162	6.3	1954	8	ADK13634
35	160	6.2	1230	2	AAW98275
36	160	6.2	1230	2	AAW98275
37	156.5	6.1	3073	6	ABU21223
38	155	6.0	270	3	AAW73072
39	155	6.0	270	3	AAW73072
40	151.5	5.9	2204	6	ABU36440
41	149.5	5.8	1063	5	ABW54168
42	149.5	5.8	1180	3	AAW01845
43	149.5	5.8	1188	3	AAW01844
44	148.5	5.7	273	2	AAW20658
45	147	5.7	1468	7	ADD48744

ALIGNMENTS

RESULT 1

AAB36552
ID AAB36552 standard; protein; 486 AA.

XX AAB36552;

XX 07-MAR-2001 (first entry)

XX Lawsonia intracellularis flagellar hook protein FlgE SEQ ID NO:1.

XX Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;
XX intestinal disease; immunogenic; diagnosis; antibacterial; swine; pig;
XX infection; detection; identification.

XX Lawsonia intracellularis.

XX WC200069904-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-AU000437.

XX 13-MAY-1999; 99US-0133973P.

XX (PFIZ) PFIZER PROD INC.
XX (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
XX (PIGR-) PIG RES & DEV CORP.

XX Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;
XX Ankenbauer RG;

XX WPI; 2001-016210/02.
XX N-PSDB; AAC86037.

XX New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody, in
XX useful in vaccines and diagnosis of Lawsonia infections, particularly in
XX swine.

XX Claim 13; Page 87-90; 97pp; English.

XX The present sequence is the Lawsonia intracellularis flagellar hook
XX protein FlgE. The present invention describes an isolated or recombinant
XX polypeptide (I) that comprises, mimics or cross-reacts with a B- or T-
XX cell epitope of a FlgE (flagellar hook) polypeptide from a Lawsonia spp.
XX (I) has antibacterial activity, and induces a specific humoral immune
XX response. (I) are used as antigens in vaccines to prevent or treat
XX infection by Lawsonia, in birds and animals, especially pigs, to raise
XX specific antibodies (Ab) and to detect past or present infection. Ab are

19-JUN-2003 (first entry)
 Protein encoded by Prokaryotic essential gene #8366.
 Antisense; prokaryotic essential gene; cell proliferation; drug design.
 Burkholderia mallei.
 W0200271783-A2.
 03-OCT-2002.
 21-MAR-2002; 2002WO-US009107.
 21-MAR-2001; 2001US-00815242.
 06-SEP-2001; 2001US-00948993.
 25-OCT-2001; 2001US-0342923P.
 08-FEB-2002; 2002US-00072851.
 06-MAR-2002; 2002US-0362699P.
 (ELIT-) ELITRA PHARM INC.
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 WPI; 2003-029926/02.
 N-PSDB; ACA26709.
 New antisense nucleic acids, useful for identifying proteins or screening
 for homologous nucleic acids required for cellular proliferation to
 isolate candidate molecules for rational drug discovery programs.
 Claim 25; SEQ ID NO 50763; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
 the 5213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
X. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
 the target prokaryotic essential genes. Note: The sequence data for this
 patent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

Sequence 413 AA;
 Query Match 19.6%; Score 506.5; DB 6; Length 413;
 Best Local Similarity 27.2%; Pred. No. 2,2e-29;
 Matches 137; Conservative 81; Mismatches 180; Indels 105; Gaps 11;
 8 GATGKTHSTGLTGNNTANNTIGYKQQVVFQDLAIGSTGSGQGNQAGMAQ 67

6 GLSGLAGASSDLVIGNNANANTVGFKGSTAQFSDMYANGVA-----SAVNNFIGITM 60
 68 VGSVETITQGAPEFGNSVTDLAIGGKGFQVLTLEDKVHYTRAGNFRFTQDGLNDPSGF 127
 61 LASVQQQFSQGITSTSTSLNVALNGNGFFQMSNGVTVTSDGTFFQDKNGYIVNSQGL 120
 128 TLMGSRISNNPNKKTETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANP 187
 121 NLMGYAANANGVINTAATVPLQA-----PTTNIAPTAT-TKITQFNL-NSQDAVPA-TTP 173
 188 YFALLESWKNGTTPPISNTSYAQPMVYDQGNHSDITVYEDGAPSTSGKTFEYLV 247
 174 F-----NYTDFTS---YNYTTSVQVDFDLGSGQNNVLFYVKSATSGQWEAYAGPAG 221
 248 MNPSEBGSAAAGTSGAGLLMSGTMTFSSNGELKNNATFTPTGSAATKDLNAMQAPLVNGL 307
 222 KTFPTDLGS-----VKPSTAGTITGTSTP----- 244
 308 POFSAFVGAGIQPLTLDFGIKSQQNMWAGAPASAAAGTGTGKLPSPMPTQSSGN--- 364
 245 -----AGVPT-----TNVQFSFSITTTGAANPON 270
 365 -----STARNGSSSTRYSODGYPQGDVLDVITITSEKLGKQKYSNSQVDFYNNIPLAR 417
 271 LTLDTGTTQYGGKNGINNLADGFASGVLTFTFSIGADGKVTGNYNSGTSTLGOIVLAN 330
 418 FTSEDLREGNNHYSATLDSGPFEGIPGTGNYKLSVNQLETNSVNDMSREYNWIIQ 477
 331 FNNPGLNVNNGVQVESAASGVQTSAPGSTNHTGLOGSALENSVNDLTSQVLKLTAAQ 390
 478 RGFQWNSKVTTADTLMOKALEL 500
 391 RNYQANAQTIKTQQTVDQTLINL 413

RESULT 7
 ADK13784
 ID ADK13784 standard; protein; 402 AA.
 XX ADK13784;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE E. coli iron transport and metabolism protein SEQ ID NO:79.
 XX
 XX Escherichia coli; iron transport and metabolism protein; iron transport;
 XX iron metabolism; antibacterial.
 XX Escherichia coli.
 XX W02004018638-A2.
 XX
 PD 04-MAR-2004.
 XX
 PF 21-AUG-2003; 2003WO-US026488.
 XX
 PF 21-AUG-2002; 2002US-0405331P.
 XX
 XX (MINU) UNIV MINNESOTA.
 XX (KAPU) KAPUR V.
 XX (GADG) GADGIL M.
 XX
 XX Kapur V, Gadgil M;
 XX
 XX WPI; 2004-238974/22.
 XX N-PSDB; ADK13720.
 XX
 XX New isolated and purified iron transport and metabolism polypeptides and
 XX encoding polynucleotides, useful in identifying potential targets for
 XX agents against pathogenic bacteria.
 XX
 PS Claim 1; SEQ ID NO 79; 185pp; English.

RESULT 10	
ADF07897	
ID	ADF07897 standard; protein; 412 AA.
XX	
XX	
AC	ADF07897;
XX	
DT	12-FEB-2004 (first entry)
XX	
DE	Bacterial polypeptide #4010.
XX	
KW	Proteus mirabilis infection; bacterial infection; antibacterial;
KW	immunostimulant.
XX	
OS	Proteus mirabilis.
XX	
PN	US6605709-B1.

AC AEM68268;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens protein sequence #1365.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX
 OS Photorhabdus luminescens.
 XX
 PN WO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-IB003040.
 XX
 PR 07-FEB-2001; 2001FR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A,
 PI Buchrieser C;
 XX
 DR WPI; 2003-148459/14.
 XX
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 2; SEQ ID NO 1365; 1205pp; French.
 CC
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX
 SQ Sequence 405 AA;
 Query Match 17.3%; Score 448.5; DB 6; Length 405;
 Best Local Similarity 27.8%; Pred. No. 4.8e-25;
 Matches 141; Conservative 66; Mismatches 174; Indels 127; Gaps 17;
 QY 10 TGMKHTSTGLTVSNANANTIGYKQOVVQDLFSDQLAIGSTGSGPNQAGVQ 69
 DB 8 SGLNAAAGNLVDVGNNAANTYGFKSTVAESDIFA-----GS-----OVGLGVKS 55
 QY 70 SVRTIFTOGAFEPGNSVTDLAIGGKGFQVTLLED-KVHYTRAGNFRFTQDGLNDPSGFT 128
 DB 56 GYNQNFKDGTPTTTNRFLDLAITQGGFFRMQGGDGIYVSRNGQFKMDENRNIINQGMK 115
 QY 129 IMG-----SRISNNPIKK-ETLEPQLDNDPTVAKS--PAKTSTALNAVNLGSDTKT 181
 DB 116 LTGYPAASVNGGAPEIQKGANPEPI-----TIQGIIMAKATTKVTWTANL-NSMHHK 167
 RESULT 11
 AEM68268
 ID ABM68268 standard; protein; 405 AA.
 XX

Breton GL;
 PI
 DR WPI: 2003-895291/82.
 DR N-PSDB; ADP03725.
 XX
 PT New Proteus mirabilis polypeptides and polynucleotides, useful as
 PT reagents for diagnosis of bacterial disease, as components of
 PT antibacterial vaccines, as targets for antibacterial drugs, or as
 PT biocontrol agents for plants.
 XX
 PS Disclosure; SEQ ID NO 8182; 870pp; English.
 XX
 CC The invention relates to new Proteus mirabilis polypeptides and
 CC polynucleotides. The invention also relates to antibodies against the
 CC polypeptides, methods for producing the polypeptides, a method of
 CC generating vaccines for immunising an individual against P. mirabilis, a
 CC method for evaluating a compound for the ability to bind a P. mirabilis
 CC polypeptide and a method for screening test compounds for anti-bacterial
 CC activity. The polypeptides and polynucleotides are useful as molecular
 CC targets for diagnosing, preventing and treating pathological conditions
 CC resulting from bacterial infection, as reagents for diagnosis of
 CC bacterial diseases, as components of antibacterial vaccines, as targets
 CC for antibacterial drugs or as bio-control agents for plants. This
 CC sequence represents a Proteus mirabilis polypeptide of the invention.
 XX
 SQ Sequence 412 AA;
 Query Match 18.4%; Score 475.5; DB 7; Length 412;
 Best Local Similarity 26.7%; Pred. No. 4.7e-27;
 Matches 136; Conservative 81; Mismatches 165; Indels 127; Gaps 15;
 QY 10 TGMKHTSTGLTVSNANANTIGYKQOVVQDLFSDQLAIGSTGSGPNQAGVQ 69
 DB 14 SGLNAAAGNLVDVGNNAANTYGFKSTVAESDIFA-----GS-----GAGLGKVS 61
 QY 70 SVRTIFTOGAFEPGNSVTDLAIGGKGFQVTLLEDK---VHYTRAGNFRFTQDGLNDPSG 126
 DB 62 GISQNFKDGSTTTNRFLDLAITQGGFFRMQGGDGIYVSRNGQFKMDENRNIINQGMK 119
 QY 127 FILMSRISN-----NPNIKKLEPIQLDNDPTVAKSPATSTALNAVNLG---DSTD 179
 DB 120 MRITGYPVQNDKKNVQKGAFTPIII-----PTDMN-ASATKMDMTVNLNAAEAID 174
 QY 180 KTQSANPYFALLSWKNGTGPSTSNVSAQPMRVYDQGNSHDITVYFDGAPSGTGS 239
 DB 175 QTHKFDP-----KON-----DSYNFSTNVTYDLSLNEHNLNLF-----VKT 213
 QY 240 KTFEVLVAMNPSDEGSAAGTSDAGLLMSGTWTFSSNG-----ELKMTAFTPTGSAT 292
 DB 214 KDNESVVAQDTTTEPAQDL-----GKLVKNDGVLDETAPKLNFTTVAKGS-- 263
 QY 293 KDLNAAQWAPLVNGLPQPSANFVGAGIPLTLDPCIKSOQNNWAGAPASAAIGTDIGKL 352
 DB 264 -----QPMDEMFN----- 272
 QY 353 PSMPFIQTSSGNSTARNSSSTRYSQDQYQDGLVDVITTSSEKLOKYSNSQVDFYN 412
 DB 273 -----SGSTQKQVAESSVKLAQNGYQAGEFTNFIPEPDGSIATYSNQSQVVGQ 323
 QY 413 IPLARFTSEDEGLRRGNHYSATLSDGGPEFLPCTSNVGLSVNQLTSTNVDMVSREWN 472
 DB 324 IALANFAPGGSLSSQGDNNWSTNGSGSPIGVAGSGVFGKLTNLALEASNVDMVSQELVN 383
 QY 473 MIIIRGFQMSKSVTTTADTLMQKALELK 501
 DB 384 MIVARQVNSNAQTIKTQDQILQLVSMR 412
 RESULT 11
 AEM68268
 ID ABM68268 standard; protein; 405 AA.
 XX


```

PR 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR N-PSDB; ABD04540.
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX PS Disclosure; SEQ ID NO 19715; 455pp; English.
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective anti-bacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABO67826-
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC segdata.uspto.gov/sequence.html
XX Sequence 269 AA;

Query Match 11.9%; Score 307; DB 7; Length 269;
Best Local Similarity 21.6%; Pred. No. 1.1e-14;
Matches 108; Conservative 46; Mismatches 93; Indels 254; Gaps 11;

QY 1 MMSLSFIGATGKMTSTGLGTYSNNIANANTIGYKQOQVFDLPSQDLAIGSTGSGQPN 60
DB 9 MLSALVSKTGSADMLTISNNLANVSTGFKRDRAEFDLLYQIRRPQGSQDS 68
QY 61 QAGMGAQVGS-VRTIFTQAFEPGNSVT-----DLAIGKGFQVTLLED-KVHYTRAGNF 113
DB 69 ELPSGLQLGTGVVRVGTOKIFTPGSLQTEOPLDMANVGRGFFQVLLPDGTVSVTRDGSF 128
QY 114 RFTQDGLNDPSGFTLMGSRISNNENIKETLEPTQLDFNDPTVAKSPAKTSTALNAVNV 173
DB 129 HUNSDQIVTSGFAL-----EPAI-----148
QY 174 LGDSTDKTQSEANPYFALLESKWKGNTPTISTSYAQPVRVYDQGNSHDITVFDGA 233
DB 149 -----148
QY 234 PSTGSKTFEYLVANVPSEDGSAASGTSAGLLMSGTMTFSSGELKMTAFTTGSATK 293
DB 149 -----VVPNE-----TOTFT-----158
QY 294 DLNMQAPLVLNGLPQFSANFVGAGIQPLTLDFGKIQOQNNWAGASAAAIGTDIGKLP 353
DB 159 -----VG-----160
QY 354 SMPITSGNSTARNGSSSTRYSODGYPOGDLVDVITITSEGLQKYSNVQVDFYNI 413
DB 161 -----QDG-----TVSVTTGNAQ-----PQVIG--NI 181
QY 414 PLARFTSEDLGRREGNNHYSATLDSGGPEFGFLPGTSNYGKLSVNQLETSNVDMREVMNV 473
DB 182 QTADFINPAGLQAIIGNLNFLETGSGAPQVGTGPGNLGLGTVAQNTLENSNVNVVEELVNM 241
QY 474 IIQGFQNSKSVTTADTML 494
DB 242 ITTORAYEMNSKVIATDQML 262

RESULT 14
ABM68266
ID ABM68266 standard; protein; 261 AA.
XX AC ABM68266;
XX 20-NOV-2003 (first entry)
XX Photorhabdus luminescens protein sequence #1363.
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
XX antibiotic; biopesticide; virulence factor; disease model; plague;
XX whooping cough.
XX Photorhabdus luminescens.
XX WO200294867-A2.
XX 28-NOV-2002.
XX 07-FEB-2002; 2002WO-IB003040.
XX 07-FEB-2001; 2001FR-00001659.
XX (INSP ) INST PASTEUR.
XX (CNRS:) CNRS CENT NAT RECH SCI.
XX Duchaud E, Tacurit S, Glaser P, Frangeul L, Kunst P, Danchin A;
XX Buchrieser C;
XX WPI; 2003-148459/14.
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX Claim 2; SEQ ID NO 1363; 1205pp; French.
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX Sequence 261 AA;

Query Match 11.4%; Score 296; DB 6; Length 261;
Best Local Similarity 20.3%; Pred. No. 7.3e-14;
Matches 104; Conservative 53; Mismatches 93; Indels 262; Gaps 8;

QY 1 MMSLSFIGATGKMTSTGLGTYSNNIANANTIGYKQOQVFDLPSQDL----AIGSTGS 56
DB 1 MRSLSWIATGLDQAQTNMDVIANLANVSTGFKRRAIFEDLLYQIRPQGMSSQT 60
QY 57 QQFN--QAGMGAQVGSVRTITFTQAFEPGNSVTDLAIGKGFQVTLLEDKVH-YTRAGNF 113
```

CC	pylori infection or to identify H. pylori polypeptide binding compounds,
CC	useful as potential H. pylori life cycle activators or inhibitors. The
CC	genomic sequence of H. pylori (ATCC 55679) was determined from
CC	overlapping contigs generated by mechanically shearing the bacterial DNA.
CC	The sequences were analysed for ORF of at least 180 nucleotides, and the
CC	predicted coding regions defined by computer evaluation. To identify
CC	likely H. pylori antigens for vaccine development, the amino acid
CC	sequences predicted from various ORF were analysed for significant
CC	homology to other known or exported membrane proteins. Having identified
CC	and determined the sequences of interest, particular regions can be
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide
CC	production, e.g. in E. coli hosts
XX	Sequence 271 AA;
SQ	Query Match 11.1%; Score 288; DB 2; Length 271;
	Best Local Similarity 21.0%; Pred. No. 3.1e-13;
	Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;
QY	1 MMGSLFIGATGMKTHSTGLGTVSNNIANANTIGYQQQVVFQDLFSQDLAIGSTGSGQPN 60
DB	10 MLRSLSYATSGLMAQQTHIDTTNNIANVNTTGFYKSRADFNDFYQAMQVAGTNTSNT 69
QY	61 -----QAGMGAQVGSVRTFTQGAPEGNSVTDLAIGKGFQVTLBD-KVHYTRAGNF 113
DB	70 LSPDGMVEGLGVRSPAITNMFSGSPKTEENLDAITGKGFQVQLPDGTAYTRSGNF 129
QY	114 RFTQDGFNDPSPGFTLMSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNN 173
DB	130 KLDEQGNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 159
QY	174 LGDSTDKTQSEANPYFALLESKWKGNGTPISTNSYSYAQPMRVYDQGNSHDITVYFDGA 233
DB	160 IG-----VDGT 165
QY	234 PSSTGSKTFEYLVAMNPSDEGSAASGTSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293
DB	166 VSVT----- 169
QY	294 DLNMQAPLVNGLPQPSANFVGAGIQPLTLDGFKSQQNMWAGAPASAAAIGTDIGKLP 353
DB	170 ----- 169
QY	354 SMWPIQTSSGNSSTARNGSSSTRYSQDGPQGLVDVVTITSEGKLQKYSNSQVVDVYNI 413
DB	170 -----QGLQTTNSVIG--QI 182
QY	414 PLARFTSEDLRRREGNNHYSATLDGGPEFGLPGTSNYGKLSVNOLETNSVDMSEWNNM 473
DB	183 TLANFVNPAGLHSMGDNLFSTINASGDAIVGNPDSQGLKRLQGFLELSNVLVEEMTDL 242
QY	474 IIRQGFQWNSKSVTTADTMLOKALELKR 502
DB	243 ITAQRAYEANSKSIQTADAMLOTVNSLKR 271
	RESULT 16
	AAW24645
ID	AAW24645 standard; protein; 271 AA.
XX	AAW24645;
AC	AAW24645;
XX	
DT	11-AUG-1997 (first entry)
XX	H. pylori flagella associated protein, 29298130.aa.
DE	
XX	Chronic gastritis; duodenal ulcer disease; activator; inhibitor;
KW	bacterial life cycle; vaccine; immunisation; detection; antisense;
KW	inhibition; flagella; flagellum; basal body; rod.
XX	
OS	Helicobacter pylori.
XX	
PN	WO9719098-A1.

DB	61 RLPGLQIGTVRPVATERIHSQNLSETNSHNSHVAIRGKGFQVQLPDGTSAYTRDGSF 120
QY	114 RFTQDGFNDPSPGFTLMSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNN 173
DB	121 QEDQNGQLTTANGFLIF-----PTI----- 140
QY	174 LGDSTDKTQSEANPYFALLESKWKGNGTPISTNSYSYAQPMRVYDQGNSHDITVYFDGA 233
DB	141 ----- 140
QY	234 PSSTGSKTFEYLVAMNPSDEGSAASGTSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293
DB	141 ----- 140
QY	354 SMWPIQTSSGNSSTARNGSSSTRYSQDGPQGLVDVVTITSEG-----KLQKYSNSQVVD 409
DB	141 -----TIPDNAT-----DLTISDGIYSVKVQSQVPOQVGQ 172
QY	410 FYNIPLARFTSEDLRRREGNNHYSATLDGGPEFGLPGTSNYGKLSVNOLETNSVDMSE 469
DB	173 F---TLTTFINDSGLESIGENLYVETSSGVPIENTPGINGAGLLQHKYVETSNVIAEE 229
QY	470 MYNMIIRQGFQWNSKSVTTADTMLOKALELK 501
DB	230 LVNMIQRAYEVNSKAVSDQMLQKLTQLZ 261
	RESULT 15
	AAW20373
ID	AAW20373 standard; protein; 271 AA.
XX	AAW20373;
AC	
DT	29-JUL-1997 (first entry)
XX	H. pylori flagella associated protein, 29298130.aa.
DE	
XX	Vaccine; prevention; treatment; infection; identification;
KW	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW	duodenal ulcer disease; chronic gastritis; diagnosis; flagella;
KW	associated; flagellum.
XX	
OS	Helicobacter pylori.
XX	
FN	WO9640893-A1.
XX	19-DEC-1996.
XX	06-JUN-1996; 96WO-US009122.
XX	07-JUN-1995; 95US-00487032.
PR	01-APR-1996; 96US-00630405.
XX	(ASTR) ASTRA AB.
PA	
XX	Smith D, Berglindh OT, Mellgaerd BL;
PI	
XX	WPI; 1997-052306/05.
XX	N-PSDB; AAT67783.
DR	
XX	Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT	useful for vaccines to treat or prevent H. pylori infection, and to
PT	detect Helicobacter.
XX	
XX	Claim 56; Page 560-561; 1481pp; English.
PS	
XX	The present sequence is a Helicobacter pylori flagella associated
CC	protein. The protein may be used in a vaccine to prevent or treat H.

XX PD 29-MAY-1997.
 XX XX
 XX PF 15-NOV-1996; 96WO-US018542.
 XX XX
 XX PR 17-NOV-1995; 95US-00561469.
 XX XX
 XX PA (ASTR) ASTRA AB.
 XX XX
 XX PI Smith DH;
 XX XX
 XX DR WPI; 1997-298052/27.
 XX DR N-PSDB; AAT77463.
 XX XX
 XX PT Helicobacter pylori nucleic acid sequences and related proteins - used
 XX PT for diagnostics and therapeutics.
 XX XX
 XX PS Claim 18; Page 167-168; 235pp; English.
 XX XX
 XX CC The present sequence is a Helicobacter pylori flagella associated
 XX CC protein, which was found to be homologous to flagellar basal body rod
 XX CC protein following BLAST protein analysis. H. pylori has been strongly
 XX CC linked to chronic gastritis and duodenal ulcer disease. The nucleic acid
 XX CC sequences of the invention are used to evaluate compounds, especially
 XX CC activators or inhibitors of bacterial life cycle, for the ability to bind
 XX CC an H. pylori nucleic acid sequence. The nucleic acid sequences, and
 XX CC corresponding proteins, are also useful for generating vaccines for
 XX CC immunising subjects against H. pylori or for use in detecting the
 XX CC presence of Helicobacter species in a sample. Antisense nucleic acid
 XX CC sequences of these sequences are used to inhibit expression of a gene
 XX CC from Helicobacter species. H. pylori whole genomic DNA was isolated and
 XX CC nebulised to a median size of 2000 bp. Purified DNA fragments were blunt-
 XX CC ended and ligated to unique BstXI-linker adapters in 100-1000 fold molar
 XX CC excess. These linkers are complementary to the BstXI-cut pMPX vectors,
 XX CC while the overhang is not self-complementary. Therefore the linkers will
 XX CC not concatamerise nor will the cut vector re-ligate itself easily. The
 XX CC linker-adaptor inserts were ligated to each of the 20 pMPX vectors to
 XX CC construct a series of shotgun subclone libraries. The purified DNA
 XX CC samples were then sequenced. Note: The ORF/protein reference number for
 XX CC this sequence was obtained from the related specification, WO9640893
 XX XX
 XX SQ Sequence 271 AA;

Query Match 11.1%; Score 288; DB 2; Length 271;
 Best Local Similarity 21.0%; Pred. No. 3, 1e-13;
 Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;

QY 1 MMGSLFIGATGKTHSTGLGTSSNNIANANTIGYKQOQVVFQDLPSQDLAIGTSGSQPN 60
 DB 10 MLRSLSYATSGMLAQOQTHIDTTSNNIANVNTGFKSRADFNDFYQAMQVAGTNTSNT 69
 QY 61 -----QAGMGAQGVSVRTITQAFEPGNSVTDLAIGKGFFOVTLDD-KVHTVRAAGN 113
 DB 70 LSPDGMVEVLGVSPSAITMYSQSPKSTENNLIAITGKGFQVQVQPDGTTATRGNF 129
 QY 114 RFTQDGFNDPSGFTLMGSRISNNPNIKETLEFIQDLDFNDPTVAKSPAKTSTALNAVYN 173
 DB 130 KLDEQGNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 159
 QY 174 LGDSDTKTQSEANPYFALLESKNGKNGTPTTSTNSVYAQPMRVYDQGNSHDINVPDGA 233
 DB 160 IG-----VDGT 165
 QY 234 PSTSGKTFEYLVAMNPSEDGSAASGSDSAGLMSGTMTFSSNGELKNMTAFTPTGSATK 293
 DB 166 VSVT----- 169
 QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTDLFGIKSQQNWAGAPASAAAIGTDIGKLP 353
 DB 170 ----- 169
 QY 354 SMMPIQTSSGNSTARNGSSSTRYSQDGYPOGDLVDVTTITSEGKLGKYSNSQVVDYFNI 413

DB 170 -----QGLQTTNSVIG--QI 182
 QY 414 PLARFTSEDLRREGNNHYSATLDSGPFGLPGTNSYKLSVNOLETNSVDMKREVMNM 473
 DB 183 TLANFVNPAGLHSMGDNLFSTWASDAIVGNPDSQGLKRGQFLELSNVRLVEMTDL 242
 QY 474 IITQRFQMNKSVTTADTMLQKALELKR 502
 DB 243 ITAQRAYEANSKSIQTADAMLTQVNSLKR 271

RESULT 17
 AAW20863
 ID AAW20863 standard; protein; 279 AA.
 XX AC AAW20863;
 XX DT 16-JUL-1997 (first entry)
 XX DE H. pylori flagella-associated membrane protein, 129e20305orf1.
 XX KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 XX KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 XX KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; flagella.
 XX OS Helicobacter pylori.
 XX PN WO9640893-A1.
 XX PD 19-DEC-1996.
 XX PF 06-JUN-1996; 96WO-US009122.
 XX PR 07-JUN-1995; 95US-00487032.
 XX PR 01-APR-1996; 96US-00630405.
 XX PA (ASTR) ASTRA AB.
 XX PI Smith D, Berglinth OT, Mellgaard BL;
 XX DR WPI; 1997-052306/05.
 XX DR N-PSDB; AAT68116.
 XX PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
 XX PT useful for vaccines to treat or prevent H. pylori infection, and to
 XX PT detect Helicobacter.
 XX PS Claim 56; Page 1266-1267; 1481pp; English.
 XX CC The present sequence is a Helicobacter pylori flagella-associated
 XX CC membrane protein. The protein may be used in a vaccine to prevent or
 XX CC treat H. pylori infection or to identify H. pylori polypeptide binding
 XX CC compounds, useful as potential H. pylori life cycle activators or
 XX CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was determined
 XX CC from overlapping contigs generated by mechanically shearing the bacterial
 XX CC DNA. The sequences were analysed for ORF of at least 180 nucleotides, and
 XX CC the predicted coding regions defined by computer evaluation. To identify
 XX CC likely H. pylori antigens for vaccine development, the amino acid
 XX CC sequences predicted from various ORF were analysed for significant
 XX CC homology to other known or exported membrane proteins. Having identified
 XX CC and determined the sequences of interest, particular regions can be
 XX CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 XX CC production, e.g. in E. coli hosts
 XX SQ Sequence 279 AA;

Query Match 11.1%; Score 288; DB 2; Length 279;
 Best Local Similarity 21.0%; Pred. No. 3, 2e-13;
 Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;

QY 1 MMGSLFIGATGKTHSTGLGTSSNNIANANTIGYKQOQVVFQDLPSQDLAIGTSGSQPN 60
 DB 18 MLRSLSYATSGMLAQOQTHIDTTSNNIANVNTGFKSRADFNDFYQAMQVAGTNTSNT 77

QY 61 -----QAGMAGVGSVRTIFTOGAPEGNSVTDLAIGGKGFQVLTLED-KVHYTRAGNF 113
Db 78 LSPDGEVLGVRPSAITKMFSGSKETENNLDAITGKGFQVQLPDGTTAYTRSGNF 137
QY 114 RFTQDGLNDPQFTLMGSRISNNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVNV 173
Db 138 KLDEQGNLVTSEGYLLI-----PQI--TL-----PEDTQ-----VN 167
QY 174 LGDSTDKTQSEANPYFALLESKWNGCTPPISTSNYSYAQPMRVYDQGNSHDITVYDGA 233
Db 168 IG-----VDGT 173
QY 234 PSSTGSKTFEYLVMNPSEDGSAASGTSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293
Db 174 VSVT-----177
QY 294 DLNAWQAPLVNGLPQFSANFVAGIQPLTLDFGIKQQNMWAGAPASAAAIGTDIGKLP 353
Db 178 -----177
QY 354 SNMPTQSSGNSTARNSSSTRYSQDGYPPQGLVDVTTITSEKLGKYSNQVDFYNI 413
Db 178 -----QGLQTSNVIG--QI 190
QY 414 PLARFTSEDLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMSEWVM 473
Db 191 TLANFVNPAGLHSGNDLFSITNASGDIAIVGNPDQGLKRGFLSNNRVLVEEMTDL 250
QY 474 IITQGFQMNKSVTVTADTMLOKALELKR 502
Db 251 ITAQRAYEANSKSIQTADAMLOTVNSLKR 279

RESULT 18
ADF07825
ID ADF07825 standard; protein; 265 AA.
XX ADF07825;
XX
XX 12-FEB-2004 (first entry)
XX Bacterial polypeptide #3938.
XX
XX Proteus mirabilis infection; bacterial infection; antibacterial;
XX immunostimulant.
XX
XX Proteus mirabilis.
XX
XX US6605709-B1.
XX
XX 12-AUG-2003.
XX
XX 05-APR-2000; 2000US-00543681.
XX
XX 09-APR-1999; 99US-0128706P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX
XX WPI; 2003-895291/82.
XX N-PSDB; ADF03653.
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
XX reagents for diagnosis of bacterial disease, as components of
XX antibacterial vaccines, as targets for antibacterial drugs, or as
XX biocontrol agents for plants.
XX
XX Disclosure; SEQ ID NO 8110; 870pp; English.
XX
XX The invention relates to new Proteus mirabilis polypeptides and
XX polynucleotides. The invention also relates to antibodies against the

CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
XX Sequence 265 AA;
Query Match 11.1%; Score 286; DB 7; Length 265;
Best Local Similarity 19.7%; Pred. No. 4.2e-13;
Matches 100; Conservative 55; Mismatches 98; Indels 254; Gaps 9;
QY 1 MWGSLFTGATGKTHSTGLGTVSNNIANANTIGYKQOVVQDFLFSQDL-----AIGSTGS 56
Db 6 MRSLLWIAKTGLDAQQNMMDVISNNLANVSNFGRQRAVDFEDLLYQTIQPGAMTSEQT 65
QY 57 QGPN--QAGMAGVGSVRTIFTOGAPEGNSVTDLAIGGKGFQVLTLEDKVH-YTRAGNF 113
Db 66 NAPSGLQITGTVRPAVERLHSGNLAQTNTNRVALKGGFFHVQLPDGTDATRDGSGF 125
QY 114 RFTQDGLNDPQFTLMGSRISNNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVNV 173
Db 126 QMDQNGQLVTSGGQIVPA-----IILPETAKKVMV-----156
QY 174 LGDSTDKTQSEANPYFALLESKWNGCTPPISTSNYSYAQPMRVYDQGNSHDITVYDGA 233
Db 157 -----GRDGIIVSEIEGS 169
QY 234 PSSTGSKTFEYLVMNPSEDGSAASGTSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293
Db 170 PA-----171
QY 294 DLNAWQAPLVNGLPQFSANFVAGIQPLTLDFGIKQQNMWAGAPASAAAIGTDIGKLP 353
Db 172 -----PQ-----QVQQL- 178
QY 354 SNMPTQSSGNSTARNSSSTRYSQDGYPPQGLVDVTTITSEKLGKYSNQVDFYNI 413
Db 179 -----TLTT-----182
QY 414 PLARFTSEDLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMSEWVM 473
Db 183 -----FINDSGLSEVGENLYLETASSCAPTENAPGINGAGLLYQGYVETSNVNAEELVNM 238
QY 474 IITQGFQMNKSVTVTADTMLOKALEL 500
Db 239 IOTQRAYEANSKSIQTSQMLQKLTQL 265

RESULT 19
AA51377
ID AA51377 standard; protein; 263 AA.
XX AA51377;
XX
XX 04-MAY-2000 (first entry)
XX
XX C. jejuni flgG protein.
XX
XX Basal body rod protein; flgG; flagellum; vaccine; immunogen;
XX pathogenic bacteria; detection; antibacterial.
XX
XX Campylobacter jejuni.
XX
XX US6020125-A.
XX
XX 01-FEB-2000.
XX

PF 07-JUN-1995; 95US-00483857.
 PR 08-MAY-1995; 95US-00436748.
 XX (CONN-) CONNAUGHT LAB LTD.
 PA Louie H, Chan VL;
 XX WPI: 2000-146875/13.
 DR N-PSDB; AA288554.
 XX Recombinant basal body rod protein producible by a transformed host for
 PT use in immunological vaccine compositions for in vivo administration to
 PT protect against diseases caused by bacterial pathogens.
 XX
 XX Example 1; Col 23-26; 28pp; English.
 XX This invention describes a novel recombinant basal body rod protein (I)
 CC producible by a transformed host containing an expression vector
 CC comprising a nucleic acid selected from; (a) an entire nucleic acid
 CC sequence of 1800 base pairs (bp), or the nucleic acid sequence of the
 CC Flg gene having 810 bp, both given in the specification; (b) a nucleic
 CC acid encoding the amino acid sequence of the flgG protein, a 270 residue
 CC sequence, given in the specification; (c) a nucleic acid sequence
 CC encoding a functional flgG basal body rod protein of a flagellum of a
 CC strain of Campylobacter; or (d) an immunogenic fragment of an flgG
 CC protein of (a), (b) or (c), and expression means operatively coupled to
 CC the nucleic acid molecule for expression by the host of a basal body rod
 CC protein of a flagellum of a strain of Campylobacter. (I) is useful in
 CC immunological vaccine compositions for in vivo administration to protect
 CC against diseases caused by bacterial pathogens that produce basal body
 CC rod proteins. (I) are also useful as immunogens, as antigens in
 CC immunoassays, or for procedures for the detection of antibacterial,
 CC Campylobacter, basal body rod protein and/or peptide antibodies. The
 CC product of the invention has immunostimulatory activity. This sequence
 CC represents the Campylobacter jejuni flgG protein described in the method
 CC of the invention
 XX Sequence 263 AA;
 Query Match 9.2%; Score 237.5; DB 3; Length 263;
 Best Local Similarity 18.6%; Pred. No. 1.8e-09;
 Matches 95; Conservative 45; Mismatches 115; Indels 255; Gaps 9;
 QY 1 MMSLFIGATGCMKTHSTGLGTVSNINANTIGYKQOQVVDLFSQDLAIGTSGQGFN 60
 DB 1 MMSLHTAATGNVAAQQQIDVTSNNIANVNTAGPKSRABFADLMYQVMKYAGTSTATT 60
 QY 61 QAGMGAQVG-----SVRTIFTOGAF- PGNSTVDLAIGKGFFQVLTLED-KVHYTRAGN 112
 DB 61 LSPSGIEVGVRPTAVTKVFTGKLSSTDTGLDMAIAGNGFFQIQLPDGTIGYTRNGQ 120
 QY 113 PRFTQDGLNDPSGFTLMGSRISNNPNIKKETLEPIQLDNDFTVAKSPAKTSTALNAV 172
 DB 121 FTKNEGVNVDGYRL----- 137
 QY 173 NLGSDTDKTOSEANPYFALLESMKNGTGPPISTSNYSYAQPMRVYDQGNSHDITVYFDG 232
 DB 138 ----- 137
 QY 233 APSSTGSKTEYLVMNPSEDGASCTDSAGLLMGSTMTFTSSNGELKNTAFTPTGSAT 292
 DB 138 ----- 137
 QY 293 KDLNMQPAPLVNGLPOFSAFVAGIQPLTDFGIKSCQNMWAGAPASAAIGTDIGKL 352
 DB 138 -----LPE-----MTIEG-----ATAINATD-GTV 158
 QY 353 PSMPIOTSSGNSTARNSSSTRYSQDGYPOGDLVDVITTSKLGQKYSQSVDFYN 412
 DB 159 SVMLP-----GEQQTQI---GQ 173
 QY 413 IPLARFTSDGLRRGNHNSATLDSGGPFGLPGTSTNYGKLSVNQKTSNVDMSREYN 472

DB 174 VELVQFINPAGLHSMGDNLYLETCASGAPVAGIAGQGLGTRHGFIELSNVQVLEMTD 233
 QY 473 MIIIQRFQWNSKSVTTADTMTLQKALELKR 502
 DB 234 LITQRAYEAGSKAITTSDMLGIVNQLKX 263
 RESULT 20
 ID AAW73073
 XX AAW73073 standard; protein; 263 AA.
 XX AAW73073;
 XX 05-JAN-1999 (first entry)
 DT XX
 DE FlgG protein.
 XX FlgG operon; flgG protein; flgG protein; basal body rod protein;
 KW flagellum; secretory diarrhoea; enteritis; vaccine; therapy.
 XX Campylobacter jejuni.
 OS Campylobacter jejuni.
 XX US5827654-A.
 PN 27-OCT-1998.
 PD 08-MAY-1995; 95US-00436748.
 PF 08-MAY-1995; 95US-00436748.
 PR 08-MAY-1995; 95US-00436748.
 XX (UTOR) UNIV TORONTO.
 PA Louie H, Chan VL;
 PI WPI: 1998-593983/50.
 XX DR N-PSDB; AAV58977.
 XX DNA encoding Campylobacter flagellum basal body rod proteins - useful for
 PT recombinant production of the proteins for use as vaccines against the
 PT bacterium, and for its detection, additionally with antibodies raised
 PT with the protein.
 XX Claim 1; Fig 1; 27pp; English.
 PS This sequence represents the Campylobacter flgG protein, encoded by the
 CC flgG operon of the invention. The flgG and flgG proteins are basal body
 CC rod proteins of the flagellum of the Campylobacter. The flgG operon and
 CC host cells containing it, are useful for the recombinant production of
 CC Campylobacter, especially C. jejuni, basal body rod proteins. This
 CC bacterium is the cause of secretory diarrhoea and enteritis. As such, the
 CC peptides produced can be used to raise antibodies, which in turn can be
 CC used to detect the presence of the organism diagnosis of the conditions.
 CC Additionally, the peptides, and specifically the live vectors (e.g. pox-
 CC or vaccinia virus) can be used as vaccines against the bacterium, and the
 CC antibodies can be used for passive immunisation. The nucleic acids can
 CC also be used to detect the presence of the bacterium
 XX Sequence 263 AA;
 Query Match 8.6%; Score 223.5; DB 2; Length 263;
 Best Local Similarity 19.3%; Pred. No. 2e-08;
 Matches 98; Conservative 43; Mismatches 115; Indels 253; Gaps 12;
 QY 1 MMSLFIGATGCMKTHSTGLGTVSNINANTIGYKQOQVVDLFSQDLAIGTSGQGFN 60
 DB 1 MMSLHTAATGNVAAQQQIDVTSNNIANVNTAGPKSRABFADLMYQVMKYAGTSTATT 60
 QY 61 QAGMGAQVG-----SVRTIFTOGAF- PGNSTVDLAIGKGFFQVLTLEDKVHYTRAGN 113
 DB 61 LSPSGIEVGVRPTAVTKVFTGKLSSTDTGLDMAIAGNGFFQIQ----- 108
 QY 114 RPTQDGLNDPSGFTLMGSRISNNPNIKKETLEPIQLDNDFTVAKSPAKTSTALNAV 173

XX PF 07-OCT-1999; 99WO-CA000938.
 XX PR 07-OCT-1998; 98US-00167568.
 XX PR 08-DEC-1998; 98US-00206942.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Loosmore SM, Yang Y, Klein MH;
 XX DR WPI: 2000-303789/26.
 XX DR N-PSDB; AA52196.
 XX PT Nucleic acid molecule for producing recombinant high molecular weight
 XX PT proteins of Haemophilus which are used as a vaccine to provide protection
 XX PT against Haemophilus induced diseases in humans.
 XX PS Claim 8; Fig 28F-Q; 307pp; English.
 XX CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked to
 CC a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
 CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMWA
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMWA. The invention also discloses hmwA genes (AA52175-AS2198)
 CC and HMWA proteins (AA501824-B01849) from the non-typeable H. influenzae
 CC strains J09C, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or cell-
 CC mediated immune response to provide protection against diseases in humans
 CC caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and
 CC tracheobronchitis). The HMW proteins are also useful as antigens in
 CC immunoassays for detecting antibodies against Haemophilus, HMW proteins
 CC and/or HMW peptides. The nucleotide sequences encoding the HMW proteins
 CC can be used to isolate and clone hmw genes from other non-typeable
 CC strains of Haemophilus via hybridisation reactions. The present sequence
 CC represents a mature HMWA protein from a non-typeable strain of H.
 CC influenzae. (Updated on 12-SEP-2003 to standardise OS field)
 XX XX Sequence 1095 AA;
 XX Query Match 6.7%; Score 173.5; DB 3; Length 1095;
 XX Best Local Similarity 22.8%; Pred. No. 0.00088;
 XX Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
 QY 12 MKHTSTGLGTVSNNIAN-TIGYKQOVVFDLFSODLAIGTSGOGPQACMGQVGS 70
 Db 570 IKAGVDGNSDSDATNNANLTIKTKEKL-----TQDLNI-----SGFNKAEITAKDGS 618
 QY 71 VRTFTQGAFFPGNS-----VTDLAIGGKGFQVLTEDKHVYTRAGNFRFTQDGLND 123
 Db 619 DLITGNTNSADGTNAKVTNFQVSKISADG-HKVTLHSEKVE--TSGSNNTEDSSDNN 675
 QY 124 PSGTLMGSRISNPNKETEPIQLDFNDPTVAKSPAKTSTALNANVNLGDSPTKQTS 183
 Db 676 -AGLTIDAKNVTNNNTTSHKAVSI-----SATSGEITTKTGTINATTGNVEITATQTS 729
 QY 184 EANPYFALLSESGKNGF-----EPISNYSYAPKRVYDQGNSHDITVYFPGAPSTG 238
 Db 730 -----ILGGIESSGSVTLTATEGALVSNIS-----GNTVTVTAN-SCALTTLA 773
 QY 239 SKTPEYLVANPNSEDSAAAGTOSAGLL---MSGTWTFSNGELKNWTFPTFGSATKDL 295
 Db 774 GSTIKGTESVTSQSGDIDGTTISGGTVEVKATESLTQTSNISKIKATTGEANVTSATGTI 833
 QY 296 NAWQAPLNVNGLPQFSANFVAGTQPLTLDFGIKSGQNMWAGAPASAAAGITDGLPSM 355

Db 834 GGTISGNTVN-----VTANAGDLTVNGAE-----INATECAATLTSSGKL----- 875
 QY 356 MPIQTSSGNTARNGSSSTRYSQDGYPOGDL--VDVTTTSEGL-----QKG----- 401
 Db 876 ----TTEASSHITSAGQVNLQAQDGSVAGSINAANVTILNTTGTTLTVKGSNINATSGTL 931
 QY 402 YNSQVVDYFNIPLARFTSEGLRREGNNHYATLDSGCPGFLPCTSNYKGLSVNQLFT 461
 Db 932 VINAKDAELNGAALGNHTVVNATNANGSGSVIATSS---RVNITG----DLITINGL-- 982
 QY 462 SNVDMRSREMNMMIIIOFGQMNSKSVTTADTNLQKALELKR 502
 Db 983 -NI-TSKNGINTVLL-KGVKIDVKYIQGIASVDEIEAKR 1020
 RESULT 23
 AAR41723
 ID AAR41723 standard; protein; 1536 AA.
 XX AC AAR41723;
 XX AC AAR41723;
 DT 25-MAR-2003 (revised)
 DT 26-APR-1994 (first entry)
 XX DE High molecular weight protein 1 (HMW1).
 XX HMW; high molecular weight protein; virus; vaccine; influenza; epitope;
 XX immunity; haemophilus influenzae.
 XX OS Haemophilus influenzae.
 XX PN W09319090-A1.
 XX PD 30-SEP-1993.
 XX PF 16-MAR-1993; 93WO-US002166.
 XX PR 16-MAR-1992; 92GB-00005704.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Barenkamp SJ;
 XX DR WPI: 1993-320683/40.
 XX DR N-PSDB; AAQ49506.
 XX PT High molecular weight surface proteins - of non-typeable haemophilus
 XX which exhibit immunogenic properties.
 XX PS Claim 3; Fig 2; 100pp; English.
 CC The isolation and purification of the high molecular weight protein
 CC enables the identification of the major protective epitopes of the
 CC protein by conventional epitope mapping. These epitopes can then be
 CC synthesised using standard techniques and incorporated into fully
 CC synthetic or recombinant vaccines. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 XX SQ Sequence 1536 AA;
 XX Query Match 6.7%; Score 173.5; DB 2; Length 1536;
 XX Best Local Similarity 22.8%; Pred. No. 0.0014;
 XX Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
 QY 12 MKHTSTGLGTVSNNIAN-TIGYKQOVVFDLFSODLAIGTSGOGPQACMGQVGS 70
 Db 1011 IKAGVDGNSDSDATNNANLTIKTKEKL-----TQDLNI-----SGFNKAEITAKDGS 1059
 QY 71 VRTFTQGAFFPGNS-----VTDLAIGGKGFQVLTEDKHVYTRAGNFRFTQDGLND 123
 Db 1060 DLITGNTNSADGTNAKVTNFQVSKISADG-HKVTLHSEKVE--TSGSNNTEDSSDNN 1116

QY 124 PSQFTLMGSRISNNPNKIKETLEPIQDFNDPTVAKSPAKTSTALNAVNLGSDTKTQS 183
DB 1117 -AGLTIDAKNVTNNNITSHKAVSI-----SATSGBITTKGTINATGNVEITAQGS 1170
QY 184 EANPYFALLESWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGT 238
DB 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTITLA 1214
QY 239 SKTFEYLVANMPSEDSAAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTTGSATKDL 295
DB 1215 GSTIKGTESVTTSSQSGDIGTISGGTVEVKATESLTTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPKM 355
DB 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNSSSTRYSQDGPQDGL---VDVTITSEGL---QK- 401
DB 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAANVTLTGTLTVKGSNINATSGTL 1372
QY 402 YSNQVDFYNIPLARFTSEDLRRGNHYSATLDSGGFGLPGTSNYKLSVNOLET 461
DB 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATTS---RVNITG---DLITINGL-- 1423
QY 462 SNVDMSEVMNMIIOQGFQNMKSQSVTTADTLMQKALELKR 502
DB 1424 -NI-ISKXGINTVLL-KGVKIDVKYIQPGIASVDEVEIAKR 1461

RESULT 24
AA41725
ID AAR41725 standard; protein; 1536 AA.

AC AAR41725;
XX
XX
DT 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX
XX High molecular weight protein 1 (HWM1).
XX
KW HWM; high molecular weight protein; virus; vaccine; influenza; epitope;
KW immunity; haemophilus influenzae; gene cluster.
XX
OS Haemophilus influenzae.
XX
XX Key Location/Qualifiers
FH Misc-difference 668..677
FT /note= "Possibly incorrect sequence. Alternative sequence
FT for this region is LNVSEGEFN. (See comments)."

XX W09319090-A1.
XX
XX 30-SEP-1993.
XX
XX 16-MAR-1993; 93WO-US002166.
XX
XX 16-MAR-1992; 92GB-00005704.
XX
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Barenkamp SJ;
PI
XX WPI; 1993-320683/40.
DR N-PSDB; AAQ49508.
XX
XX High molecular weight surface proteins - of non-typeable haemophilus
PT which exhibit immunogenic properties.
XX
XX Claim 3; Figure 2/10; 100pp; English.
XX
XX The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be

CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines. This sequence is claimed to be the
CC same as that given in AAR41723 (High molecular weight protein 1) although
CC it does differ slightly. (Repeated regions which are possibly incorrect
CC and occur in the corresponding nucleotide coding sequence contribute to
CC these differences). (Updated on 25-MAR-2003 to correct FN field.)
XX
XX Sequence 1536 AA;

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 0.0014;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MHTSTGLGTVSNNIANAN-TIGYKQOQVFDLFSQDLAIGTSGSQPNQAGGAQVGS 70
DB 1011 IKAGVDGSDSDATNNANLTIKTELK-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTFTQCAFPGNS-----VTDLAIGGKFFQVLEDKVHVTRAGNPRFTQDGLND 123
DB 1060 DLTIGNTNSADGTNAKVTFFNQKDKLSADG-HKVTILHSKE--TSGSNNTEDSDNN 1116
QY 124 PSQFTLMGSRISNNPNKIKETLEPIQDFNDPTVAKSPAKTSTALNAVNLGSDTKTQS 183
DB 1117 -AGLTIDAKNVTNNNITSHKAVSI-----SATSGBITTKGTINATGNVEITAQGS 1170
QY 184 EANPYFALLESWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGT 238
DB 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTITLA 1214
QY 239 SKTFEYLVANMPSEDSAAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTTGSATKDL 295
DB 1215 GSTIKGTESVTTSSQSGDIGTISGGTVEVKATESLTTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPKM 355
DB 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNSSSTRYSQDGPQDGL---VDVTITSEGL---QK- 401
DB 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAANVTLTGTLTVKGSNINATSGTL 1372
QY 402 YSNQVDFYNIPLARFTSEDLRRGNHYSATLDSGGFGLPGTSNYKLSVNOLET 461
DB 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATTS---RVNITG---DLITINGL-- 1423

RESULT 25
AAR63505
ID AAR63505 standard; protein; 1536 AA.

XX AAR63505;
XX
XX 25-MAR-2003 (revised)
DT 25-JUN-1995 (first entry)
XX
XX Haemophilus high molecular weight protein HWM1.
DE High molecular weight protein; HWM1; protective vaccine; otitis;
XX sinusitis; bronchitis; Hib.
XX Haemophilus.
XX
XX WO9421290-A1.
XX
XX 29-SEP-1994.
XX
XX 15-MAR-1994; 94WO-US002550.
XX
XX 16-MAR-1993; 93US-00038682.
PR

XX PA (BARE/) BARENKAMP S J.
XX PA (SGEN/) ST GEME J W.
XX PI Barenkamp SJ, St Geme JW;
XX DR WPI: 1994-316665/39.
XX DR N-PSDB; AAQ72293.
XX PT New immunogenic high mol. wt. proteins of non typeable Haemophilus -
XX PT useful in protective vaccines.
XX PS Claim 2; Page 31; 127pp; English.
XX CC The HmW1 protein encoded by this sequence is useful in a vaccine to
XX CC protect against disease caused by non-typeable Haemophilus which are not
XX CC controlled by H. influenzae type b (Hib) vaccines. The encoded protein
XX CC can also be used as a carrier for protective Hib polysaccharide (in a
XX CC conjugate vaccine against meningitis) or for other antigens, haptens, etc.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 1536 AA;
Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 0.0014;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTHSTGLTVSNINAN-TIGYKQOVVFQDLFSQDLAIGTSGQPNQAGNGAQVGS 70
Db 1011 IKAGVDGENSDATNNANLTIKTKEKL-----TQDLNI-----SGFNKAETAKDGS 1059
QY 71 VRTIFQGAPEFGNS-----VTDLAIGGKFFQVTLKDVHYTRAGNFRFTDQGLND 123
Db 1060 DLTIGNTNSADGTNAKVTNFQKDSKISADG-HKVTLSHKE--TSGSNNTIEDSSDNN 1116
QY 124 PSGFLMGRISNPNKETEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQS 183
Db 1117 -AGLIDAKNVTYNNITSHKAVSI-----SATSGEITTKTGTINATTGNVEITAGTGS 1170
QY 184 EANPYFALLEKWKNGT-----PISTSNYSYAQPMRVYDQGNSHDITVYFGAPSGTG 238
Db 1171 ----ILGGIESSGSVTLTATEGALVSNIS-----GNTVTYAN-SGALTTLA 1214
QY 239 SKTEYLVAMPSPEDGSAAGTDSAGLL---MSTWTFSSNGELKNMTPTPGSATKDL 295
Db 1215 GSTIKGTESVTTSSQSDIGTISGGTVEVKATESLTQSNKIKATGTGEANVTSGTI 1274
QY 296 NAWQPAPLVNLGPFQSANFVCGAGIOTLDFGIKSQNMWAGAPASAAAGTIDIGKLP 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATITTSGLK--- 1316
QY 356 MPIOTSSGNSTARNSSSTRYSODYPOQDL---VDVITTSSEKL---QSK----- 401
Db 1317 ----TTEASSHITKAGQVNLSDAGSVAGSINAANVTTLNTTGLTVKGSNINATSGTL 1372
QY 402 YSNQVVDVFNPIARFTSEDLRREGNNHYSATLDSGGPEFGIPGTFNSYKLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVNAVNGSGVIATSS---RVNITG---DLITINGL--- 1423
QY 462 SNVDMSEVMNMIIQRFQNMNSKSVTTADTMLOKALELKR 502
Db 1424 -NI-ISRNGINTVLL-KGVKIDVKYIQFGIASVDEIVEAKR 1461
RESULT 26
AAB01846
ID AAB01846 standard; protein; 1536 AA.
XX AC AAB01846;
XX DT 12-SEP-2003 (revised)
XX DT 11-SEP-2000 (first entry)

DE XX Haemophilus influenzae strain 12 HmW1A protein, SEQ ID NO:67.
XX KW HmW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
XX KW recombinant production; Escherichia coli; antibacterial; vaccine;
XX KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX KW detection; diagnosis.
XX OS Haemophilus influenzae; strain 12.
XX PN WO2000020609-A2.
XX PD 13-APR-2000.
XX PF 07-OCT-1999; 99WO-CA000938.
XX PR 07-OCT-1998; 98US-00167568.
XX PR 08-DEC-1998; 98US-00206942.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Loosmore SM, Yang Y, Klein MH;
XX WIPI; 2000-303789/26.
XX DR N-PSDB; AAA52195.
XX PT Nucleic acid molecule for producing recombinant high molecular weight
XX PT proteins of Haemophilus which are used as a vaccine to provide protection
XX PT against Haemophilus induced diseases in humans.
XX PS Example 16; Fig 28A-Q; 307pp; English.
XX CC The invention relates to the recombinant production of Haemophilus
XX CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX CC expression construct used to effect recombinant expression comprises a
XX CC promoter functional in E. coli (e.g., the 17 promoter) operably linked to
XX CC a modified hmwABC operon from a non-typeable (non-encapsulated) H.
XX CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
XX CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
XX CC and the hmwB and hmwC genes encode accessory proteins which are
XX CC responsible for post-translational processing and secretion of the HMW
XX CC proteins. The modified hmwABC operon used in the expression construct of
XX CC the invention contains an A gene modified such that it encodes only the
XX CC mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
XX CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
XX CC strains Jcy, K1, K21, LDCD2, PMH1, 15 and 12. The nucleic acids and
XX CC vectors are used for the production of recombinant H. influenzae HMW
XX CC proteins which can be used as vaccines to mediate a humoral or cell-
XX CC mediated immune response to provide protection against diseases in humans
XX CC caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and
XX CC tracheobronchitis). The HMW proteins are also useful as antigens in
XX CC immunoassays for detecting antibodies against Haemophilus, HMW proteins
XX CC and/or HMW peptides. The nucleotide sequences encoding the HMW proteins
XX CC can be used to isolate and clone hmw genes from other non-typeable
XX CC strains of Haemophilus via hybridisation reactions. The present sequence
XX CC represents an HMW protein from a non-typeable strain of H. influenzae.
XX CC (Updated on 12-SEP-2003 to standardise OS field)
XX SQ Sequence 1536 AA;

Query Match 6.7%; Score 173.5; DB 3; Length 1536;
Best Local Similarity 22.8%; Pred. No. 0.0014;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTHSTGLTVSNINAN-TIGYKQOVVFQDLFSQDLAIGTSGQPNQAGNGAQVGS 70
Db 1011 IKAGVDGENSDATNNANLTIKTKEKL-----TQDLNI-----SGFNKAETAKDGS 1059
QY 71 VRTIFQGAPEFGNS-----VTDLAIGGKFFQVTLKDVHYTRAGNFRFTDQGLND 123
Db 1060 DLTIGNTNSADGTNAKVTNFQKDSKISADG-HKVTLSHKE--TSGSNNTIEDSSDNN 1116

QY 124 PSQFTLMGRISNPNKIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQS 183
DQ 1117 -AGLTIDAKNVTVNNITSHKAVSI-----SATSGEITTKTGTTINATGNVEITAQTS 1170
QY 184 EANPYFALLESWKNGT-----PPISTSNYSVAQPMRVYDQGNSHDIIVYFDGAPSSTG 238
DQ 1171 ----ILGGIESSGSVLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVANPSEDSAGSTDSAGLL---MSGTWFSSNGELKNTAFTPTGSAKDL 295
DQ 1215 GSTIKGTESVTTSSQSDIGITSGTVEVKATESLTTQSNKIKATGEANVTSATGTI 1274
QY 296 NAMQAPLVNGLPQFSANFVGAGIQPLTLDFGLKSOQNMWAGAPASAAIGTDIGKLPKM 355
DQ 1275 GGTISGNTVN-----VTNAGDLTVNGAE-----INATEGAALVTTSSGKL--- 1316
QY 356 MPIQTSSGNSTARNGSSSTRYQDGYPOGDL--VDVTTTSEKL---QCK----- 401
DQ 1317 ----TTEASHITSAGQVNLQAQDSVAGSINAANVTTLTGTATVKGSNINATSGTL 1372
QY 402 YNSQVVDVFNIPARFTSEDLRRGNHYSATLDSGGPEFGLPGTSNKGKLVNOLET 461
DQ 1373 VINAKDAELGAALGNHVTNATNANGSGSVIATTS---RVNITG---DLITINGL--- 1423
QY 462 SNVDMSEVMNIIIOGFQWNSKSVTTADTMLQKALEKR 502
DQ 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQGIASVDEVIEAKR 1461

RESULT 27
ABU50020
ID ABU50020 standard; protein; 3705 AA.
AC ABU50020;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #35547.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Yersinia pestis.
XX
FN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
FA (ELIT-) ELITRA PHARM INC.
XX
FI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
XX
DR N-PSDB; ACA53890.
XX

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.
PS Claim 25; SEQ ID NO 77944; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway or a gene on which the test compound that inhibits proliferation of an identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a product's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3705 AA;
QY Query Match 6.7%; Score 172.5; DB 6; Length 3705;
DQ Best Local Similarity 21.7%; Pred. No. 0.0059;
QY Matches 119; Conservative 58; Mismatches 180; Indels 191; Gaps 24;
DQ 3 GSLFIGATCMKTHSTGLGTSVNNIANANTIGYKQQVVFQDLFSODLAIGTSGSGPNOA 62
DQ 2359 GELIIGANGAFQOTSLNIAAG--ASANINGYRQ-----TVGAVNTGTVTTL 2403
QY 63 GMAQVGSVRIFTQGAPEGNSVTDLAIGKGPQVTLKDVHYTRAGNFRFTODGFLN 122
DQ 2404 GNGELTGTDLTINTGMINVTDGILNLENGGASSISGGL-----TNGILN 2449
QY 123 DPSSG-FTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKT 181
DQ 2450 IKGGDFTI-----SIDNN-----GLAQTNISDGASVT 2477
QY 182 QSEANPYFALLESWKNGTTPPISTSNYSVAQPMRVYDQGNSHDIIVYFDGAPSSGT 241
DQ 2478 -----LNGGTTIIGTN-----LGSSV 2494
QY 242 FEYLVANPSEDSAGS-----GT--DSAGLIMSGTWTFS-----SNGEL 283
DQ 2495 IDVLGDLNLVADNSLANVISGDTTNTATVTLSSNSSFSGAHQIQTGELTVGQASNLG 2554
QY 284 AFTPTGSATKOLNAPQAPLVNGLPQFSANFVGAGIQPLTLD----- 325
DQ 2555 A---SSATVNLGTLTSLHLNMGVSESIAN-VLSGVAGSTVDIIGGADTALTANNSGFLG 2609
QY 326 -----FGIKSOQNMWAGAPASAAIGTDIGKLPMPQTSSGNSTARNGSSSTR 376
DQ 2610 QVALAGNSKLTVASTNNL--GASSVALAGT--GDTLSGNGTFTGNSVTSGLVQVT- 2664
QY 377 YSQDGYPOGDLVDVTTTSEKLQKGYNSQVVDVFNIPL-----ARFT---SEDGLRRG 428
DQ 2665 -----DDAEVLTSSNGV---SNAVTTIDADATLNLDDIALFNHALTGGLNLVA 2711
QY 429 NNHYSATLDSGGPEFG-LPGTSNKGKLVNOLETNSVDMSEVMNIIIOGFQWNSKSV 487
DQ 2712 KNDASTAFDFGATVGGAFTGTVN-----LNNSTFDLSGNNTT-VLAQATLKUSSGNL 2762
QY 488 TTADTMLQ 495
DQ 2763 TSVGNQVQ 2770

the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 5.4%; Score 166; DB 6; Length 2468; Best Local Similarity 21.9%; Pred. No. 0.01; Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;

20 GTVNNIANANTIGYKQQQVWFQDLFSDLAIGSTGSGQ-----PNOAGMGAQVGSV 71
1298 GTVNAVA-----QDPA-GNTGPGQSTTVDAVAPNTVPVNSGNL 1337
72 RTITQGAFFPGNSVTDLAIGGKGFQVTLKDVHYTRAGNFRFTQDGFNDPSGFTLMG 131
1338 ----LNGTAEPGSTVTLTDGNGNPIGQTADG-----SGNWSFTPGSLPNGTVNVTA 1387
132 SRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTOSEANPYFAL 191
1388 SDAAGNTSLPAT-----TVDSLSIPQVDPNSGVSIGTADAGNTIIITDNGNPIGQV 1443
192 LESWKNG--TPPISTSNYSYAQPMRVYDQGNSHDITVYFDG-APSGTSGKTFEYLIVAM 248
1444 TADGSGNWSFTPGIPLDGTVVNVVARSFNSVDSAPAVITVDGVAPAAP-----VI 1494
249 NPSEGGNAAGTDSAGLLM-----SGTWTF-----SNGELKNMTAFT 286
1495 DPS-NGTEISGTAEAGATVILTDGGNPIGQATADGSGNWTFTPTSTPLANGTVINVAQD 1553
287 PTG-----SATKDLNAWQAPLVN-----GLPQFSANFV-----GAGIPLTIDFG 327
1554 PAGNTSGPASVTVDAIA-PPAPVNPNSGVVISGTAEAGATVILTDGNGNPIGQVTA-- 1610
328 IKSQNNWAGAPASAAIGTDIGKL-----PSMFIQTSSGNSTA 367
1611 ---GSGNWSFTPGTPLANGSVINALAQDAAGNNSPTSATVDSLAPAPVIDPNSGVIA 1667
368 RNGSSSTRYSQD--GYPOGLVDVTITSEGLQKYSNSQVDFYNIPLARFTSEDLR 425
1668 GTAEGATVILTDGNGNPIGQ-----VTADG--SGNWSFTPGTPLSNGTVNVVAQDA-- 1718
426 REGNNH--YSATLDSGGPFGPLGTSTNYGKLS 455
1719 -AGNTSGVSTTVDAVAPATPVIDPNSGVNELS 1749

RESULT 30
ABP59933
ID ABP59933 standard; protein; 2468 AA.
XX
AC ABP59933;
XX
DT 28-AUG-2003 (first entry)
XX
DE Microbial resistance gene PA1874 protein.
XX
KW Biofilm; microbial resistance.
XX
OS Unidentified.
XX
PN WO2003041483-A2.
XX
PD 22-MAY-2003.
XX
PF 18-SEP-2002; 2002WO-US029565.
XX
PR 18-SEP-2001; 2001US-0323241P.
XX
PA (DART-) DARTMOUTH COLLEGE.
XX
PI O'toole GA, Mah T;
XX
WP; 2003-468567/44.
DR N-PSDB; ACC59398.
XX
PT Identifying modulators of microbial resistance of organisms in biofilms, e.g. inhibitor of biofilm formation, by employing expression controls, or efflux pumps containing polypeptides, of genes associated with biofilm resistance.
XX
PS Claim 33; Fig 7; 102pp; English.
XX
CC The present invention relates to a method of identifying a compound capable of altering the sensitivity of a microorganism to an antimicrobial agent by employing efflux pumps comprising polypeptides encoded by the following genes: PA1874, PA4142, PA2389, PA1876, PA4143, PA2390 or PA1163. The method is useful for identifying modulators of microbial resistance of an organism in a biofilm. The methods are also useful for identifying genes that encode proteins that play a role in biofilm resistance. The method is particularly useful for screening compounds or discovering compositions that will inhibit biofilm formation and overcome their resistance mechanisms. These methods are particularly useful in medical, industrial or natural settings, where formation of biofilms can have serious negative consequences and result in high costs both in human health and economic terms. The present sequence is a protein shown in the exemplification of the invention

Query Match 6.4%; Score 166; DB 6; Length 2468; Best Local Similarity 21.9%; Pred. No. 0.01; Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;

20 GTVNNIANANTIGYKQQQVWFQDLFSDLAIGSTGSGQ-----PNOAGMGAQVGSV 71
1298 GTVNAVA-----QDPA-GNTGPGQSTTVDAVAPNTVPVNSGNL 1337
72 RTITQGAFFPGNSVTDLAIGGKGFQVTLKDVHYTRAGNFRFTQDGFNDPSGFTLMG 131
1338 ----LNGTAEPGSTVTLTDGNGNPIGQTADG-----SGNWSFTPGSLPNGTVNVTA 1387
132 SRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTOSEANPYFAL 191
1388 SDAAGNTSLPAT-----TVDSLSIPQVDPNSGVSIGTADAGNTIIITDNGNPIGQV 1443
192 LESWKNG--TPPISTSNYSYAQPMRVYDQGNSHDITVYFDG-APSGTSGKTFEYLIVAM 248
1444 TADGSGNWSFTPGIPLDGTVVNVVARSFNSVDSAPAVITVDGVAPAAP-----VI 1494

RESULT 34
 ADK13634
 ID ADK13634 standard; protein; 1954 AA.
 AC ADK13634;
 DT 17-JUN-2004 (first entry)
 DE L. kirschneri bacterial-Ig-like (Big) domain containing protein BigL2.
 KW gene therapy; vaccine; infection; Leptospira; bacterial-Ig-like;
 KW Big domain; BigL2.
 OS Leptospira kirschneri.
 FH Key Location/Qualifiers
 FT Misc-difference 338
 FT /note= "Encoded by AGTA"
 XX US2004058323-A1.
 XX 25-MAR-2004.
 XX 19-SEP-2002; 2002US-00147299.
 XX 19-SEP-2002; 2002US-00147299.
 XX (KQAI/) KO A I.
 XX (REIS/) REIS M G.
 XX (CROD/) CRODA J H.
 XX (SIQU/) SIQUEIRA I C.
 XX (MATS/) MATSUNAGA J.
 XX (RILE/) RILEY L W.
 XX (BARO/) BAROCCHI M A.
 XX (YOUN/) YOUNG T A.
 XX Ko AI, Reis MG, Croda JH, Siqueira IC, Matsunaga J, Riley LW;
 XX Barocchi MA, Young TA;
 XX WPI; 2004-268783/25.
 XX N-PSDB; ADK13633.
 XX New substantially purified BigL1, BigL2 and BigL3 polypeptides and
 XX encoding polynucleotides, useful for diagnosing, preventing or treating
 XX Leptospiral infections in humans and other mammals.
 XX Claim 4; SEQ ID NO 4; 52pp; English.
 XX The invention relates to a substantially purified polypeptide. The
 XX methods and compositions of the present invention are useful for the
 XX diagnosis, prevention and/or treatment of infection with Leptospira
 XX species in humans and other mammals, including those of veterinary
 XX importance. The present sequence represents the amino acid sequence of
 XX the Leptospira kirschneri bacterial-Ig-like (Big) domain containing
 XX protein BigL2.
 XX Sequence 1954 AA;
 Query Match 6.3%; Score 162; DB 8; Length 1954;
 Best Local Similarity 19.3%; Pred. No. 0.015;
 Matches 125; Conservative 80; Mismatches 241; Indels 200; Gaps 23;
 7 ICATGMKTHSTGLG-TVSNNIYANA-----NTIGYKQQQVVFQDLFSQDL 49
 410 VGSNTIAAIGVDITVSLNVTNATLESIQVSDSHSIARGTSTFVQAIYVSDGSSONI 469
 50 -----AIGTSGSQGNQAGMGQVGSVRTIFTQGAFFPGNSVTDL 89
 470 SDQVANNSSNIIQIENLNNAVPKREIQSPSSGGLG-----TARTATLEIAISSYTDI 522
 90 AIGKGFFQVTLDEKVVHTRAG-NFRFTQDGLNDPSGFTMGSRISNNPNKIKETL--- 145

Db 523 SYNAATLVSEIVSPNTPSVSSGLTVPFATGTYTGGNQNLTSQVTWNSNTNRATISNA 582
 Qy 146 ---EPIQLDFNDPT-----VAKSPAKTSTALNAVVN-----LGDSTDKT- 181
 Db 583 NGTQIALGSSVGTNNISATLGAVTSSATTLTVNAVLSNITIPSLPSVAVGSLNLTA 642
 Qy 182 ----OSEANPYFALLESWKNGTGPPISTNSVSYAQMRYVYDQGNSH----- 224
 Db 643 TGTYSNGSNQDLTTSVAMTSTDSISVSDNASGROGQTTGVAQGNTOISATLGSTSSAIN 702
 Qy 225 -----DITVYFDGAPSSGTSTFEVLVAMNPSEDSAAAGTDSA----- 263
 Db 703 FTVSAVLDSIOVTLEDSPFIAGTST--RAATGVFSDGSNLSNLSQVINDSSQTNVQL 760
 Qy 264 GLLMSG---TWTFSSNGELKMTAFTPTGSAT-----KDLNAWQPA----- 301
 Db 761 GVLETPKKKLMNSPANG---NSTTGTSRITATLGGVSGYADLTIVAPSLTSIQIDPHTP 817
 Qy 302 PLVNGLPQESANFVAGIQTLDGFIKSOQNM-----WAGAPASAAAGTIDIGKLPWM 356
 Db 818 SVANGLTQ---NFTATGV-----YSDGSNQNLTDSTVWASSNPAVATISNAGTNGKAT 868
 Qy 357 PIQTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGLQKGYNSQVDFYNIPLA 416
 Db 869 TLQTGSTNISASLGATTS-----DPSVLTVT-----NATLTSITIAPTS 907
 Qy 417 RFTSEDGLRRE-----GNHYSATLDSGGPFGPLPGTSNYGKL 454
 Db 908 SENIAGLNDQFVATGYTGGSRDLTTQVTWNSNTSTATISNANGTQGRMAAVDTGST 967
 Qy 455 SVN-----QLETSNVDSREVMANMIIQRFQFMNSKSVTTADTWL 494
 Db 968 NISASLGGTYSQTTNVTTSVAVLSIQVSPADISVAKGNTKAYTAI 1013
 RESULT 35
 AAW98275
 ID AAW98275 standard; protein; 1230 AA.
 XX AAW98275;
 AC AAW98275;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPO 690 protein.
 XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 XX peptic ulcer disease.
 XX Helicobacter pylori.
 XX WO9843478-A1.
 XX 08-OCT-1998.
 XX 01-APR-1998; 98WO-US006371.
 XX 01-APR-1997; 97US-00833457.
 XX 24-JUN-1997; 97US-00891227.
 XX 29-JUL-1997; 97US-00902615.
 XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 XX (HUNA-) HUMAN GENOME SCI INC.
 XX Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;
 XX WPI; 1998-542293/46.
 XX N-PSDB; AAX13994.
 XX New isolated Helicobacter polynucleotides - used to develop products for
 XX the diagnosis, prevention and treatment of Helicobacter infections and
 XX gastrointestinal diseases.

```
PS Claim 8; Page 301-306; 2054pp; English.
XX This sequence represents a Helicobacter pylori CHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastrointestinal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis
XX
SQ Sequence 1230 AA;
Query Match 6.2%; Score 160; DB 2; Length 1230;
Best Local Similarity 21.6%; Pred. No. 0.011;
Matches 114; Conservative 71; Mismatches 182; Indels 160; Gaps 25;
QY 20 GTVSNNIANANTIGYKQQQV--FQDLFSQDLAIGS--TGSQGNQAGMGAQVGSVRTTFT 76
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
542 GTTTNTQAKSNAPKLKAMVVVNEEEAKTANLAQSGSTTTQSPNSTVMGA----LNTVL- 596
QY 77 QGAFEPGNSVTDLAIGKGFFQVLTLEDKVHYTRAGNFRFT-QDGFNDPDSGFTLMGSRIS 135
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 -----QNVSNFQOSIQNAFQNESNIQAWANAY 625
QY 136 NNPNIKKETLEP-----IQDLFNDPTVAKS--PAKTSTALNAVNLGSDTDKQSEANP 187
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 NTNGSQSQEMTPNNQDLRIQLRANFYQLINTINQOVPTDMNALINQSQOQTQSGSASN 685
QY 188 YFALLESKNGKNGTPTISTNSYSAQPMRVYDQGNSHDITVYFDGAPSGTSGKTFEYLVA 247
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
686 NNACASGMSG-----SNGWCYQQ-----WSDSKAYYSGLSALGYQT----- 723
QY 248 MNPSEDGSAAGTDSAGLLMSGT--MTFSSNGELK-----NMTAFTPTGSATK 293
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
724 -----QATTQSGSGNGNSIYVQOITLTNGLNLQIITNLKSVNGNGASGTSGNGHTS 778
QY 294 DLNAWQAPLVNGLPQPSANFVAGIQPLTLDGFIKQQNMWAGAPASAAAGTIDGKLP 353
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
779 QIN-----TAYQMLT-----DASDGKLGST----- 797
QY 354 SMWPIQTSSG-----NSTARGSSSTR-RYSQDGYPOGLVDVVTITSEGLQCKYSN 404
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
798 -----YSSSGSGNGYTPCNSIT--NSGKTSNNCYEPKQONATTATATDSNLQKVYND 851
QY 405 SQVVDYFNIPLARFTS---EDGLRR--EGNNHYSATLDS--GGPEPGLPQTSNYGKL--S 455
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
852 AQKI--ANIIASSGNNKGVNGLKQPFPEALKNNSSLSNLKNGSGSGSGSTTCGWLINL 909
QY 456 VNQLETNSVMSREWNMI---IIQRCGF-QWNSKSVTTADTMLQKAL 498
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
910 LGAIPTNGVSDTNLLNLTETFIKTAGFIQNNSSSVSTSLTSAFOAI 956
```

```
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0036299P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX N-PSDB; ACA34839.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 58993; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1230 AA;
Query Match 6.2%; Score 160; DB 6; Length 1230;
Best Local Similarity 21.6%; Pred. No. 0.011;
Matches 114; Conservative 71; Mismatches 182; Indels 160; Gaps 25;
QY 20 GTVSNNIANANTIGYKQQQV--FQDLFSQDLAIGS--TGSQGNQAGMGAQVGSVRTTFT 76
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
542 GTTTNTQAKSNAPKLKAMVVVNEEEAKTANLAQSGSTTTQSPNSTVMGA----LNTVL- 596
QY 77 QGAFEPGNSVTDLAIGKGFFQVLTLEDKVHYTRAGNFRFT-QDGFNDPDSGFTLMGSRIS 135
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 -----QNVSNFQOSIQNAFQNESNIQAWANAY 625
QY 136 NNPNIKKETLEP-----IQDLFNDPTVAKS--PAKTSTALNAVNLGSDTDKQSEANP 187
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 NTNGSQSQEMTPNNQDLRIQLRANFYQLINTINQOVPTDMNALINQSQOQTQSGSASN 685
QY 188 YFALLESKNGKNGTPTISTNSYSAQPMRVYDQGNSHDITVYFDGAPSGTSGKTFEYLVA 247
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
686 NNACASGMSG-----SNGWCYQQ-----WSDSKAYYSGLSALGYQT----- 723
```


AAW73072

ID AAW73072 standard; protein; 270 AA.

XX AC AAW73072;

XX DT 05-JAN-1999 (first entry)

XX DE FlgF protein.

XX KW FlgFG operon; flgF protein; flgG protein; basal body rod protein; flagellum; secretory diarrhoea; enteritis; vaccine; therapy.

XX OS Campylobacter jejuni.

XX PN US5827654-A.

XX PD 27-OCT-1998.

XX PF 08-MAY-1995; 95US-00436748.

XX PR 08-MAY-1995; 95US-00436748.

XX PA (UTOR) UNIV TORONTO.

XX PI Louie H, Chan VL;

XX DR WPI; 1998-593983/50.

XX DR N-PSDB; AAV58977.

XX PT DNA encoding Campylobacter flagellum basal body rod proteins - useful for recombinant production of the proteins for use as vaccines against the bacterium, and for its detection, additionally with antibodies raised with the protein.

XX PS Claim 1; Fig 1; 27pp; English.

XX CC This sequence represents the Campylobacter flgF protein, encoded by the flgFG operon of the invention. The flgF and flgG proteins are basal body rod proteins of the flagellum of the Campylobacter. The flgFG operon and host cells containing it, are useful for the recombinant production of Campylobacter, especially C. jejuni, basal body rod proteins. This bacterium is the cause of secretory diarrhoea and enteritis. As such, the peptides produced can be used to raise antibodies, which in turn can be used to detect the presence of the organism diagnosis of the conditions. Additionally, the peptides, and specifically the live vectors (e.g. phage or vaccinia virus) can be used as vaccines against the bacterium, and the antibodies can be used for passive immunisation. The nucleic acids can also be used to detect the presence of the bacterium

XX SQ Sequence 270 AA;

Query Match 6.0%; Score 155; DB 2; Length 270;

Best Local Similarity 30.7%; Pred. No. 0.003; Matches 46; Conservative 28; Mismatches 64; Indels 12; Gaps 5;

QY 1 MMGSLFYGATGKMTHTSLGTGTVSNNTANANTIGYKQOVV---FQDLFSQ---DLAIGS- 53

Db 1 MONGYYQATGGMVTFQFKLDVITNNLANINTSGYKDDVVIADFKRIFKETQDELPIENH 60

QY 54 --TGSQGNQAGMG-AQVGSVRTIFTQGAPEPNSVTDLAIGKGPFQV--TLEDKVHYT 108

Db 61 TRDASRFVNTTIDGIPQVQSQYTFDLSGLSKATNPLDLAMTREDAFYLVQTKDGEVRLT 120

QY 109 RAGNFRFTQDGLNDPSGFTLMGSRISNPN 138

Db 121 KQGNFQLDDEGLVKNQGYKVLSSDYFNNP 150

RESULT 39

AAV51376

ID AAV51376 standard; protein; 270 AA.

XX AC AAV51376;

XX DT

04-MAY-2000 (first entry)

XX DE C. jejuni flgF protein.

XX KW Basal body rod protein; flgF; flagellum; vaccine; immunogen; pathogenic bacteria; detection; antibacterial.

XX OS Campylobacter jejuni.

XX PN US6020125-A.

XX PD 01-FEB-2000.

XX PF 07-JUN-1995; 95US-00483857.

XX PR 08-MAY-1995; 95US-00436748.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Louie H, Chan VL;

XX DR WPI; 2000-146875/13.

XX DR N-PSDB; AAZ88553.

XX PT Recombinant basal body rod protein producible by a transformed host for use in immunological vaccine compositions for in vivo administration to protect against diseases caused by bacterial pathogens.

XX PS Claim 1a; Col 21-24; 28pp; English.

XX CC This invention describes a novel recombinant basal body rod protein (I) producible by a transformed host containing an expression vector comprising a nucleic acid selected from: (a) an entire nucleic acid sequence of 1800 base pairs (bp), or the nucleic acid sequence of the flgF gene having 810 bp, both given in the specification; (b) a nucleic acid encoding the amino acid sequence of the flgF protein; a 270 residue sequence, given in the specification; (c) a nucleic acid sequence encoding a functional flgF basal body rod protein of a flagellum of a strain of Campylobacter; or (d) an immunogenic fragment of an flgF protein of (a), (b) or (c), and expression means operatively coupled to the nucleic acid molecule for expression by the host of a basal body rod protein of a flagellum of a strain of Campylobacter. (I) is useful in immunological vaccine compositions for in vivo administration to protect against diseases caused by bacterial pathogens that produce basal body rod proteins. (I) are also useful as immunogens, as antigens in immunoassays, or for procedures for the detection of antibodies. The Campylobacter, basal body rod protein and/or peptide antibodies. The product of the invention has immunostimulatory activity. This sequence represents the Campylobacter jejuni flgF protein described in the method of the invention

XX SQ Sequence 270 AA;

Query Match 6.0%; Score 155; DB 3; Length 270;

Best Local Similarity 30.7%; Pred. No. 0.003; Matches 46; Conservative 28; Mismatches 64; Indels 12; Gaps 5;

QY 1 MMGSLFYGATGKMTHTSLGTGTVSNNTANANTIGYKQOVV---FQDLFSQ---DLAIGS- 53

Db 1 MONGYYQATGGMVTFQFKLDVITNNLANINTSGYKDDVVIADFKRIFKETQDELPIENH 60

QY 54 --TGSQGNQAGMG-AQVGSVRTIFTQGAPEPNSVTDLAIGKGPFQV--TLEDKVHYT 108

Db 61 TRDASRFVNTTIDGIPQVQSQYTFDLSGLSKATNPLDLAMTREDAFYLVQTKDGEVRLT 120

QY 109 RAGNFRFTQDGLNDPSGFTLMGSRISNPN 138

Db 121 KQGNFQLDDEGLVKNQGYKVLSSDYFNNP 150

RESULT 40

ABU36440

ID ABU36440 standard; protein; 2204 AA.
 AC ABU36440;
 XX
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by Prokaryotic essential gene #21967.
 DE
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 XX Mycobacterium tuberculosis.
 OS
 XX WO200277183-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 XX
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 PI
 DR WPI: 2003-029926/02.
 DR N-PSDB; ACA40310.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 25; SEQ ID NO 64364; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: the sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 2204 AA;

Query Match

5.9%; Score 151.5; DB 6; Length 2204;

Best Local Similarity 20.2%; Pred. No. 0.11;
 Matches 119; Conservative 64; Mismatches 220; Indels 185; Gaps 24;
 QY 3 GSLFIGATGKTHSTGLGTVSNANANTIGYKQOQVWFQDLFSODLAIGSTG----- 55
 DB 1458 GNIGIGLTG--TQIGFGSFSNHNIGLFSNGDGNVGFNSGTGNVGLGNTGANTFGIA 1515
 QY 56 SQPNQAGHQAQVGSVRTTFTQGAPEPGNSVTDLAIGGK-----GFFQV----- 99
 DB 1516 NSGSFNTGLG-NTGSTNT-----GLFNPNGVNTGVGNTGINTGSGTNTGSGT 1570
 QY 100 TLGD-----KVHYTRAGNRF-----TQDGF----- 120
 DB 1571 NLGDHNTGFSNGDYNVTFNAGDYNVNTGNTGNTGNTGNTGNTGNTGNTGNTGNTG 1630
 QY 121 -----LNDPSGFTLMGSRISNNPN-----IKKETLEP 147
 DB 1631 LSTTITIPBIPYRDLSEVIDIPITGTVVATTTPNSFTIPGQIRVLLGPAALVNMIGP 1690
 QY 148 IOLDENDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPYFALLESKNGTPTPISN 207
 DB 1691 ITIDNVQVIAIDSPIQCTLSMVGTGGFGPIPIGIGIGTGGP-----GNST----- 1736
 QY 208 YSYAOPMRYVDQGNSHDITVYFDGAPSTSTGSKTPEYLAMNPSBDGSAASG-TDSAGLL 266
 DB 1737 ---TGPSSGFFHTGAGHVSFGNFCAGNMSGNF-----GAGNSGFFNAGGLG 1782
 QY 267 MSGTWTFS-NGELKNMTAFTPTGSATKDLANQWAPLIVGLPQFSANFVGAGIQ---PL 322
 DB 1783 NSGLNFGALQSLANLGNNTISGVYNTTDLATPA-FGSGIANIGANLAGLFDNTGNL 1841
 QY 323 TLDFGIKQQNMWAGAPASAAAIGTDIGKLPMMPIQTSSGNSTARNSSSTRY----- 377
 DB 1842 TLNFGVANQGLNAG-----IGNLGSVNIGFVNTGDSNLGILNGLDNLFGGWN 1890
 QY 378 ---SQDGYQGDLVDVTITSEKLOKQYKN--SQVDF-----YNIPLARFTSD----- 422
 DB 1891 GGNNGIANTGIFDGLANLGSYNIGLANLGDNDLGFNAGSYNIGFANFGSDNLGFANT 1950
 QY 423 -----GLRREGNNHYS-----ATLDSGGEFGL--PGTSNYG 452
 DB 1951 GSYNIGFANTGNNNIGVGLTGNQIGIGLSLNSGNNIGLFSNGSGNIG 1998

Search completed: October 26, 2004, 09:05:38
 Job time : 81 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2004, 20:00:04 ; Search time 484 Seconds
(without alignments)
5318.298 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMGSLFVIGATGKTHSTGLG.....NEKSVTTADTMLQKALELKR 502

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3413475 seqs, 2563800928 residues

Total number of hits satisfying chosen parameters: 6826950

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-CG=1/USPTO spooll/US10009823/runat 26102004 100131 1718/app_query.fasta_1.647
-DB=Published Applications NA -QFMT=fastp -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10009823 @CGN 1 1 354 @runat 26102004 100131 1718
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -YGAPOF=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09C_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2586	100.0	1509	15	US-10-009-823A-2
2	529.5	20.5	1242	16	US-10-282-1222A-11957
3	527.5	20.4	1389	15	US-10-127-032-16
4	506.5	19.6	1239	16	US-10-282-1222A-14579
5	485	18.8	2157	16	US-10-335-977-58
6	484.5	18.7	1163020	16	US-10-398-221-10
7	484.5	18.7	3011208	16	US-10-398-221-2058
8	469.5	18.2	5998	16	US-10-398-221-3855
9	414.5	16.0	640681	9	US-09-790-988-1
10	296	11.4	665	16	US-10-335-977-56
11	288	11.1	789	16	US-10-335-977-1
12	288	11.1	816	16	US-10-335-977-2
13	288	11.1	837	16	US-10-335-977-3
14	237	9.2	555	9	US-09-974-300-1888
15	215.5	8.3	1524	9	US-09-974-300-1846
16	201.5	7.8	459	16	US-10-335-977-78
17	193	7.5	5445	14	US-10-210-296-2
18	193	7.5	5445	15	US-10-449-462-2
19	189.5	7.3	1395	16	US-10-335-977-82
20	180.5	7.0	1965	15	US-10-369-493-28400
21	180.5	7.0	1965	15	US-10-369-493-31160
22	173.5	6.7	3285	15	US-10-193-764-64
23	173.5	6.7	5116	13	US-10-092-880-1
24	173.5	6.7	5116	15	US-10-193-764-62
25	173.5	6.7	9171	13	US-10-092-880-5
26	172.5	6.7	11118	16	US-10-282-1222A-41760
27	166	6.4	273	16	US-10-335-977-55
28	166	6.4	7407	15	US-10-246-330-3
29	166	6.4	7407	16	US-10-282-1222A-30151
30	165.5	6.4	372	16	US-10-335-977-77
31	165	6.4	25165	14	US-10-114-170-39
32	164	6.3	1029	16	US-10-335-977-57
33	162	6.3	5863	16	US-10-147-299A-3
34	160	6.2	3693	16	US-10-282-1222A-22709
35	160	6.2	3793	9	US-09-881-752A-149
36	156.5	6.1	9219	16	US-10-282-1222A-12963
37	153.5	5.9	2037	15	US-10-369-493-37512
38	152.5	5.9	554	16	US-10-398-221-1330
39	151.5	5.9	810	16	US-10-335-977-17
40	151.5	5.9	6615	16	US-10-282-1222A-28180
41	150	5.8	3396	16	US-10-193-764-60
42	149.5	5.8	3543	15	US-10-193-764-58
43	149.5	5.8	3568	15	US-10-193-764-58
44	148.5	5.7	819	16	US-10-335-977-18
45	147.5	5.7	558	16	US-10-335-977-79

ALIGNMENTS

RESULT 1

US-10-009-823A-2
; Sequence 2, Appli
; Publication No. US20030157120A1
; GENERAL INFORMATION:
; APPLICANT: Panaccio, Michael
; APPLICANT: Rosey, Everett Lee
; APPLICANT: Sinistaj, Meri
; APPLICANT: Hasse, Detlef
; APPLICANT: Parsons, Jim
; APPLICANT: Ankenbauer, Robert G.
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE
; TITLE OF INVENTION: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES
; FILE REFERENCE: DAVI150.001APC
; CURRENT APPLICATION NUMBER: US/10/009,823A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/AU00/00437
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/133,973
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1509

; TYPE: DNA

; ORGANISM: Lawsonia intracellularis

US-10-009-823A-2

Alignment Scores:

Pred. No.:	1,05e-259	Length:	1509
Score:	2586.00	Matches:	502
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-10-009-823A-1 (1-502) x US-10-009-823A-2 (1-1509)

QY	1	MetMetGlySerLeuPheThrGlyAlaThrGlyMetLeuThrHisSerThrGlyLeuGly	20
DB	1	ATGATGGGAGTTTGTATTGTCGCAACAGTATGAAACCCCATAGTACAGGGTTGGGT	60
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal	40
DB	61	ACTGTCTCCAAATAATATTGCTAAACGCAATACCATTTGGTATATAGCAGCAAGTATG	120
QY	41	PheGlnAspPheSerClnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn	60
DB	121	TTTCAAGACCTTGTGTAGTCAAGATTAGCAATAGGTTCTACTGGAAGTCAGGGGCCAAAC	180
QY	61	GlnAlaGlyMetGlyValaGlnValGlySerValArgThrIlePheThrGlnGlyValaPhe	80
DB	181	CAGGCTGTATGGGAGCACAGGTTGGAGTGTTCGCACATTTTTCACAGGGTGTCTTT	240
QY	81	GluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPheGlnValThr	100
DB	241	GAACCTGGCAATAGTGTAAACAGATCTTGTATTTGGTGGAAAGGTTTTTTCAGGTTACA	300
QY	101	LeuGluAspLysValHisTyrArgAlaGlyAsnPheArgPheThrGlnAspGlyPhe	120
DB	301	TTAGAGGTAAGTACACTATACAGCAGCAGGAGGAAATTTTGTCTTACTCAAGATGGTTTT	360
QY	121	LeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIle	140
DB	361	TTAAATGATCTAGCGGATTTACTTTTAAATGGGCTCAAGATATCTAATAATCCTAACATA	420
QY	141	LysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer	160
DB	421	AAAAGGAAACCTTGAAACCCANTTCACTTGTAGCTTTAATGATCTCAGTAGCAAGTCT	480
QY	161	ProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLys	180
DB	481	CCTGCAAAACCAAGTACAGCAATTAACGCTGTGGTAAACCTTGGTGATAGTACAGATAAA	540
QY	181	ThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr	200
DB	541	ACAAAGTGAAGTAACTATCCATATCTTTCACCTTCTTGAGAGCTGGAAAGAAATGGAACA	600
QY	201	ProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln	220
DB	601	CCTCCTATTCTACATCAAACTACTATATGCAACACTATGAGATGATATGATCAACAA	660
QY	221	GlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLys	240
DB	661	GGAAATTTCTCAGGATATACTGTATATTTTCAATGGGAGCACCTCTTCAACAGGAAGTAAA	720
QY	241	ThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaLysGlyThr	260
DB	721	ACATTTGAATACCTTGTAGCTATGATGATGATGATGATGATGATGATGATGATGATGAT	780
QY	261	AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys	280
DB	781	GATAGTGCAGGCTCTCTTAAATGTCTGGAACACTATGACATTTTCAAGTAAATGCGCAATTA	840

QY	281	AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro	300
DB	841	AATATGACAGCTTTTACTCTCTACTGCTCTGCAACAAAGATTAAATGATGCGCAACA	900
QY	301	AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLeuGln	320
DB	901	GCACCAATTAGTCAATGGTTTACCACAGTTTTCAGCAAAATTTTGTGGTGCAGGAATACAG	960
QY	321	ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla	340
DB	961	CTTTTAACTTAGACTTTTGGAAATTAAGCCCAACAGATATGTGGCAGAGCTCCAGCA	1020
QY	341	SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetProIleGlnThr	360
DB	1021	TCGGTCTGCTCCATAGGTACAGATATTGGGAATTTGCCATCAATGATGCAATACAAACA	1080
QY	361	SerSerClyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp	380
DB	1081	TCAGCGGTAATTTCTACAGCAAGAAATGGATCATCTTCAACAAGAGATATAGCCAGAT	1140
QY	381	GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly	400
DB	1141	GGTTATCCTCAGGAGATCTAGTACATGTCACAAATTAATTTTCCCTTACGACGCTTACAAGT	1200
QY	401	LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer	420
DB	1201	AAGTATAGTAATAGTACAGTCTGTTGATTTTATAATATTTTCCCTTACGACGCTTACAAGT	1260
QY	421	GluAspGlyLeuArgArgGluGlyValAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly	440
DB	1261	GAGGATGGATTAAGACGAGAGAGGATTAACCATTTATTCGGCAACACTTGTACTCAGTGGG	1320
QY	441	ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu	460
DB	1321	CCAGAGTTTGGATTCGCAAGACATCTAACTATGGAATACTTAGTGTGAATCAACTTGAG	1380
QY	461	ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe	480
DB	1381	ACTTCTACGTAGCATGAGCAGAGAAATGGTTAATATGATTTATTTCACAGTGGTTTT	1440
QY	481	GlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu	500
DB	1441	CAGATGAATAGTAATCTGTTACACAGCAGACACAATGCTACAAAAAGCACTTGAACATA	1500
QY	501	LysArg 502	
DB	1501	AAGCGT 1506	

RESULT 2

US-10-282-122A-11957

; Sequence 11957, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR FILING DATE: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 60/206,848

; PRIOR FILING DATE: 2000-05-23

;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 11957
;; LENGTH: 1242
;; TYPE: DNA
;; ORGANISM: Burkholderia cepacia

US-10-282-122A-11957

Alignment Scores:
Pred. No.: 3,14e-45 Length: 1242
Score: 529.50 Matches: 149
Percent Similarity: 44.20% Conservative: 72
Best Local Similarity: 29.80% Mismatches: 182
Query Match: 20.48% Indels: 97
DB: 16 Gaps: 13

US-10-009-823A-1 (1-502) x US-10-282-122A-11957 (1-1242)

Qy 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAla 27
Db 16 GGCTTGAGCGATGGCGGTGGCTCGAGCAATCTCGACGTATCGGCAACATCGCG 75
Qy 28 AsnAlaAsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSergln 47
Db 76 AACGCAACACGGTCGCTTCAACGACGAGCGCGCAACTTCGCGACATGTACGCGAAT 135
Qy 48 AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln 67
Db 136 TCGGTCCGACGTCGCTCAAC-----ACGACGATCGGCATCGGACGCGG 180
Qy 68 ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr 87
Db 181 CTCGCTCGGTGACGAGAAATTCGCGCAGGAGCAGATCAATTCGACGAAATGCTGCTC 240
Qy 88 AspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyr 107
Db 241 GACGTCGCGATCAACGCGCAACGGCTTCTCCAGATGTCACGCAACGCGCGTGACCGATC 300
Qy 108 ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe 127
Db 301 TCGCGCGACGACAGTTCATCGGCAAGAACGCGCGCATGCTCGACGCGCAGGCGCG 360
Qy 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluPro 147
Db 361 AACCTGATGGCTATCG 420
Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
Db 421 CTCGACGGCG-----CCGACCAACAACTCGCGCGCGCGCGCGCGCGCGCGCGCG 465
Qy 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
Db 466 ATCCCGCGCGAGTCAACTCG-----AACGCGGACGACGAGGTGCGCGCGCAAG----- 513
Qy 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheIleSerThrSerAsn 207

Db 514 -----ACGCCGTTCAACGGCAGCGACCAAC 537
Qy 208 -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp 225
Db 538 ACGACGTACAACACTCACTCGATCCAGGTCTACGACAGCTCGCGGGCTCGCAGCAG 597
Qy 226 IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu 245
Db 598 GTCACGATGTACTTC-----GCGAAGAGCGCGCGCGCAGC----- 633
Qy 246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265
Db 634 -----TGGCTGGCCCTACGCGCGCGCTGACGCGCGCGCGCGCGCGCGCGCG 669
Qy 266 LeuMetSerGlyThrMetPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe 285
Db 670 ACGAATCTCGGACCGCTCAGCTTCGATCGCTCGGCGCGGATCAGCTCGACGACGCTCG 726
Qy 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
Db 727 GCGCGACCGCG-----CAGCCGACCGCG----- 750
Qy 306 GlyLeuProGlnPheSerAlaAsnPheVal-----GlyAlaGlyIleGlnPro 321
Db 751 AGCTTCGCGCATGTCGCGTCTTCGATCCGACACGACGCGCGCGCGCAATCCGCGAAG 810
Qy 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSer 341
Db 811 CTGACGCTCGACCTG----- 825
Qy 342 AlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSer 361
Db 825 ----- 825
Qy 362 SerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGly 381
Db 826 ---GGCGGACGACGACGACGACGCGCGCAAGGACGCGCTGACCAATCTCGCGCAGCAGCGC 882
Qy 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 883 TTCGCGACGCGCACGCTGACGAGCTTCGATCGGCGCGCGCGCGCGCGCGCGCGCGCG 942
Qy 402 TyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGlu 421
Db 943 TACTCGACGCGCAGACGCGCGGTGCTCGCGCTGATCGCGCTCGCGCAACTTCAACACCGCG 1002
Qy 422 AspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyPro 441
Db 1003 AACGCGCTCGTGAACATCGCGCGCAACCAAGTATCGGAAACGCGCGCGCTCGCGCGCG 1062
Qy 442 GluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThr 461
Db 1063 CAGATCG 1122
Qy 462 SerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGln 481
Db 1123 TCGAAGCTGAACCTGACCGACCGAGCTCGTGAACCTGATCACCAGCGCGCGCGCGCGCG 1182
Qy 482 MetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 1183 GCGAAGCGCGCAGCATCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242

RESULT 3
US-10-127-032-16
; Sequence 16 Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Bangera, M. Gita
; APPLICANT: Lory, Stephen
; APPLICANT: Greenberg, Everett Peter
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
; TITLE OF INVENTION: BIOFILM FORMATION

FILE REFERENCE: UIZ-070CP
 CURRENT APPLICATION NUMBER: US/10/127,032
 CURRENT FILING DATE: 2002-04-19
 PRIOR APPLICATION NUMBER: US 60/285,190
 PRIOR FILING DATE: 2001-04-20
 PRIOR APPLICATION NUMBER: US 60/344,142
 PRIOR FILING DATE: 2001-10-24
 NUMBER OF SEQ ID NOS: 170
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 16
 LENGTH: 1389
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-10-127-032-16

Alignment Scores:

Pred. No.: 5,97e-45 Length: 1389
 Score: 527.50 Matches: 151
 Percent Similarity: 45.84% Conservative: 86
 Best Local Similarity: 29.21% Mismatches: 205
 Query Match: 20.40% Indels: 75
 DB: 15 Gaps: 16

US-10-009-823A-1 (1-502) x US-10-127-032-16 (1-1389)

```

; 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
; 4 AGTTTCAACATCGGCTTCAGCGGCATCCAGCGGCCTTCAGCGGCCTGAACGTCAACGGC 63
; 24 AsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
; 64 AACACATCGCAACGCGCGCACCGTAGCTTCAAGCAATCCCGCGGAGTTCCGCCGAC 123
; 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlnGlyProAsnGlnAlaGly 63
; 124 GTCTACGCGCGCTCGGTG---CTGGGTTCG-----GGCAGCAACCGCGAGGC 168
; 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
; 169 AGCGCGTGTCTCTCGGAGTCTCCGAGATGTTCAAGCAGGCGCAACATCGACTCGACC 228
; 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPheGlnValThrLeuGluAsp 103
; 229 AACAGCGTGTGAGCTCGGCTCGCATCAACGCGCAACGGCTTCTTCGTCAACCAACACGGG 288
; 104 LysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
; 289 GCGATCAGCTACCCGCGCGGCTACTTCAATACCGACAGCAGGATTTTCATCGTCGAC 348
; 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
; 349 AACACAGGCTACCGCTCGGCGTATCGCTCGGCGGCAAGCGCCAGTTCGAGAACGGC 408
; 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
; 409 GTGGTACCGACCTCAAGTTCAGCGGCGCCCACTCAGGCGCGGAGGCGCACTCGAGATC 468
; 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrLysThrGln 182
; 469 CAGCGTGTGTAACACCTCAACTCGACGCTGAAG-----CCG 504
; 183 SerGluAlaAsnProTyrPheAlaLeuLeuSerTrpLysGlyAsnGlyThrProPro 202
; 505 CCGACCGTGCAGCGCTTC-----GATCCG 528
; 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsn 222
; 529 TCCGACGCGGTACTACACTCGCTCTCTCGTGGGCATCTATGACAGCCAGGCGCAAC 588
; 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
; 589 TCCACACACCATGAGCCAGTTCTTTCATCAAGAACGAGCGCGGACCGGATGCGACCCGCGG 648

```

```

; 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
; 649 ATTCCGGAGAACAGCTGGACCATGAAGTGCTGATCGACGGGTCAATCCGGTCTCATCGG 708
; 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
; 709 TCGAACACAGACGCG-----ATGAGCTTCAACGTCACCTTC 744
; 274 SerSerAsnGlyLysLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
; 745 GAGCGGCGGCGCAGATGACCTCGGTTCGGGG-----CCGACGGCAGCAGCGGG 798
; 291 -----AlaThrLysAspLeuAsnAlaThrGlnProAlaProLeuVal 304
; 799 CCGGGCTTCAGCATCGACGCGACCAACAGTGTCCAGTTCAGTCCCGGCC----- 849
; 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
; 850 ACTGGCAATCCGCGGACTCCC-----GGACCGGCTGGATTCCGGCGGCTCG 897
; 325 AspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAla 344
; 898 GAC---GGCAAGACCCCGCGGACCTACGCTGGAACGCGCGACCGGTGCGCGCCAGCGC 954
; 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
; 955 ATCTCTTCGACATCGGCAAG-----ACACCCAGTAC 987
; 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGln 384
; 988 TCCACCGCGTTCGCCGACAGCAACCCGATC-----CAGACGGCTACACACC 1035
; 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
; 1036 GTCTAGTTGGCAGGCTGGAATCGACACCGGGGTGATCTTCGCCGCTACACCAAC 1095
; 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
; 1096 GGCCAGTCCCAAGTCCAGGCGCAGGTGTGTGGCCAACTTCGCCCAACATCCAGGGCTG 1155
; 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
; 1156 AGCCCGATCGGCAAGACCTCTCTGGGTGAGTCTCTCGGAGTCCGGCGAGCCGGCGTGGC 1215
; 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
; 1216 GCGCGCGCTCGGCGACCTCGGGCGGTTCGAATCCGGCGGCTGGAGCGTCCACCGTG 1275
; 465 AspMetSerArgGluMetValAsnMetIleIleLeuGlnArgGlyPheGlnMetAsnSer 484
; 1276 GACATCTCCACGAACACTGTGAACCTCATCTCCACAGCGCAACTACAGCGCAATGCC 1335
; 485 LysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
; 1336 AAGACATCCAGACCGAGATGCGGTGACCCAGCATCATCACTTCGCGC 1386

```

RESULT 4

US-10-282-122A-14579
 ; Sequence 14579, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14579
LENGTH: 1239
TYPE: DNA
ORGANISM: Burkholderia mallei
US-10-282-122A-14579

Alignment Scores:
Pred. No.: 7,87e-43 Length: 1239
Score: 506.50 Matches: 137
Percent Similarity: 43.34% Conservative: 81
Best Local Similarity: 27.24% Mismatches: 180
Query Match: 19.59% Indels: 105
DB: 16 Gaps: 11
US-10-009-823A-1 (1-502) x US-10-282-122A-14579 (1-1239)

Qy 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla 27
Db 16 GTTTGAGCGGTTTGGCGGGCGCTCGAGCGACCTCGACGTATCGGCAACATCGCG 75
Qy 28 AsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGln 47
Db 76 AACGGAAACACGGTGGGCTTCAAGGCGACGACCGCGCAGTTTCCGACATGATCGCAAT 135
Qy 48 AspleuAlaIleGlySerThrGlySerGlnGlnProAsnGlnAlaGlyMetGlyAlaGln 67
Db 136 TCGTTCGCA-----TCGCGCGCTCAACAACTCGATCGGCATCGGCACGATG 180
Qy 68 ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr 87
Db 181 CTCGCATCGGTGACGACGACGTTGACGAGCGGCGACGATCCTCGAGCACGTCGTCGCTG 240
Qy 88 AspleuAlaIleGlyGlyGlyPhePheGlnValThrLeuGluAspLysValHisTyr 107
Db 241 AACGTCGGATCAACGGCAACGGCTTCTTCAGATGTCGAACAAACGCGCGTACCAACGATC 300
Qy 108 ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe 127
Db 301 TCGCGCGACGCGACGTTCCAGCGGCGACGAGAACGGCTACATCGTCANTTCGAGGGCGCTG 360
Qy 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluPro 147
Db 361 AACGTGATGGCTACGCGCGAAGCGGAAACGGCGTGTATCAACACCGCGGCGACCGTGC 420
Qy 148 IleGlnLeuAspPheAsnAspThrValAlaLysSerProAlaLysThrSerThrAla 167

Db 421 CTGACGCGC-----CCGACGACCAACATCGCGCCACCGCGGACG---ACCAG 465
Qy 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
Db 466 ATCACCAGCGCGCTCACTG---AACTCGAGGACGCGGTGCGCGC---ACGACGCG 519
Qy 188 TyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProPheIleSerThrSerAsn 207
Db 520 TTC-----AACTACACCGCATCCGACGACG--- 543
Qy 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227
Db 544 TAACTACTACGACGCTCGGTGCGAGGTGTTCACACGCTCGCGGCTCGCAGAACGTGAAC 603
Qy 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
Db 604 CTGTACTTCTGTAAGAGCGCGACGACGCGCGCTATGCGGGAGCGTATGCGGGCGCGCGC 663
Qy 248 MetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet 267
Db 664 AAGACGCGACCGATCTCGGCTCG----- 687
Qy 268 SerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrPro 287
Db 688 -----GTCAAGTTTCAGCACCGCGCGGACGATC 714
Qy 288 ThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeu 307
Db 715 ACGGCGACGTCGACGCGC----- 732
Qy 308 ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly 327
Db 732 ----- 732
Qy 328 IleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThr 347
Db 733 -----GCCGCGTGCAGC-----ACG 750
Qy 348 AspileGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn----- 364
Db 751 AACGTGCGCGCTTCTGTTCTCGATCCGACGACGCGCGCGCGCGCGCGCGCGCGCGCG 810
Qy 365 -----SerThrAlaArgAsnGlySerSerSerThrArgArgTyr 377
Db 811 CTGACGCTCGACCTCAGCGGCGACGACGACGACGCGCGCGCGCGCGCGCGCGCGCGCG 870
Qy 378 SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLys 397
Db 871 GCGCAGGACGCGCTTCGAGCGCGCTGCTGACGACGCTTCTCGATCGCGCGCGCGCGCG 930
Qy 398 LeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArg 417
Db 931 GTCACGCGCACTACTCGAACCGCCAGACCTCCACGCTCGCGCGCGCGCGCGCGCGCG 990
Qy 418 PheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAsp 437
Db 991 TTCAACAAACCGCGCGCTCGTGAACGTCGCGCAACACCACTAGTACGTGAGTCCGCGCG 1050
Qy 438 SerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsn 457
Db 1051 TCGGGCGTCCGACAGATCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110
Qy 458 GlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGln 477
Db 1111 GCGCTCGAAGACTCGAACGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1170
Qy 478 ArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAla 497
Db 1171 CGCACTACCG 1230
Qy 498 LeuGluLeu 500
Db 1231 ATCAACCTG 1239


```
QY 325 AspPheGly-----IleYsSerGlnGln 332
Db 1246 CAATACGGCATGTTGAAATCAACAATAAAGATAATAAATAATGCTCATTAAGAABAATCTC 1305
QY 333 AsnMetTrp----- 335
Db 1306 AATATCTTTGAGTGGGTATTCTTCAGACAGCGTAAACAAATGTTTGTATAAAGAAC 1365
QY 336 -----AlaGlyAlaProAlaSerAlaAla 343
Db 1366 GCCATGAAGAGTCTTAATACCGCTTCTTTAATTGAAGGGGAGGCGTCAGTAGCAGTTCT 1425
QY 344 -----AlaIleGlyThr--- 347
Db 1426 AAATTCAACCAACGCTACCCATGCGCAAGCAAGCATGATGTATAGACAGATTTAGGCACATAA 1485
QY 347 ----- 347
Db 1486 CACGCCATGCGCATGAGTTTATAGAGTGGGGGCGGGAATGGAATTTTAGAGTGATC 1545
QY 347 ----- 347
Db 1546 GTGCTGAGCTGGGGAATTAGTAGGGGGGTACAGCGGCTAGGCGCTAATGTGTTTGAAGGG 1605
QY 348 -----AspIleGlyLysLeuProSerMetMetPro--- 357
Db 1606 GGCGGTTTGCAATTTAATAACGACGATCGCTTCGCGGCATGAACCGCGCTCTTTTGCAA 1665
QY 358 -----IleGlnThrSerSerGlyAsnSer 365
Db 1666 TTTGACCCCTAAATGGCGTGATGCCCCCAACGCAATCAATTAGCTTTTGGTCTCTCA 1725
QY 366 ThrAlaArgAnGlySerSerThrArgArgTyrSer-----Gln 379
Db 1726 GGGAGCTTTGACGATTCACAAGCGTGGATAAGATTTCTGAAACTTATGCGATTGAGCAA 1785
QY 380 AspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGln 399
Db 1786 AACGGCTATACGGCGGATTTGATGATGATGCGCTTTCGCTTTCATCTGATGGGGTCTTTTA 1845
QY 400 GlyLysTyrSerAnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThr 419
Db 1846 GGGCGGTTTCAATGATGACGAGACTTTAGCGCTCGCTCAAGTGGCTTTAGCGAATTCGCT 1905
QY 420 SerCluAspGlyLysArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGly 439
Db 1906 AACGATCGGCGCTTACAGCGCTTTGGAGGCAATGCTCTTTTCTCAACCGGAATCTCGG 1965
QY 440 GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db 1966 CAACCCCTTAATCGTGGCGCTAATACGGGGGTAGGGTTCAATTTCAAGGATCCAAACTA 2025
QY 460 GluThrSerAnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGly 479
Db 2026 GAGTCTAGTATGTGGATTTGAGCCGAGTTTACGAAATTTGATTTGTTGTTCAAGGGGG 2085
QY 480 PheGlnMetAnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGlu 499
Db 2086 TTTCAAGCGAACTCTAAGCGGTAAACACATCCGATCCGATCAATCCTTTACACCCCTATTGAAT 2145
QY 500 LeuLysArg 502
Db 2146 CTCAGCAA 2154
RESULT 6
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
```

```
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1163020
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10
Alignment Scores:
Pred. No.: 2,8e-36 Length: 1163020
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 16 Gaps: 11
US-10-009-823A-1 (1-502) x US-10-398-221-10 (1-1163020)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 639538 ATGAATCAACATCATGTATATACAGCTATTTCTGGATGAATGGTTCCTCAACAGCATATCA 639597
QY 21 ThrValSerAsnAlaIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 639598 GTGACATCAATTAATATTCCTCAATCGCAACAGCAGCGGTATATAAACAAGCGTCGT 639657
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db 639658 TTCAATGATTTACTTTACCAAAACACAATGGATCTGTTGCGGGCGACTTTATGCTGGA 639717
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 639718 ACAAAACCAATAGAGCTTCGGTTCGGTTTCGAAATTTGGGCAATTTTAAACCGATTATACA 639777
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
Db 639778 GCAGTTCTCCGACATCACTCGCGAGAAACAAGATGAGCACTGCAAGGCCGCGCTTT 639837
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
Db 639838 TTCAATGCTGGCGATAACGCTGCTGGGAATATCGTTTACACAGCAGACGATGCTTTGCA 639897
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db 639898 GTATCTGCAATAACTATTTTAAACAGCAACAGGAAATACGTTATGGGATATGCAACA 639957
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 639958 GATAAAATATGGCAACGCTTTTAAACGGAACACTTGCAACCAATTCAAATCCCAATTAATAGC 640017
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 640018 GCAATTCAGGCGAGCAACAAAAAATGATGATTAAGCGGTAACTCCATTCCTGATTTGG 640077
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
Db 640078 GCGCAAAAA---GATACGATTTCTCCGAG----- 640104
QY 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 640105 -----CTT 640107
```

QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
Db 640108 TCTGTATACGTAATGCTGGTGGAAACATATAACTTCAAGTCAATATGAAGCTGGACA 640167
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 640168 CCAGATGGCGAGTGGCAATGTTCTCTATGAATGAATCAATG-----GAC 640215
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 640216 GGAAGCGTTAACTCTCTCA-----GTAACAGGACACCTTAATATAC 640257
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 640258 AACGGCGAAGCGCACTTACAAACCTGATGCA----- 640290
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 640291 ---CTTAAACAACTTCAATCAATCCACTGTTAACCGC----- 640326
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 640327 -----AAACAGTCAAT 640338
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 640339 ATG-----GGCTTAAACCTTAAGTGGCTTAACA 640365
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaAsnGlySerSerSer 373
Db 640366 RACTAC-----GGTACCAACCAA 640383
QY 374 ThrArgA-gTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 640384 GTATTCTCACCAACTTCTGACGGAAGCGCTGCAACTGTAAAGACTACGCACTTACC 640443
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 640444 GATTCGTGTATATGCACTGAGTTACTACAGCGTACAGTTATCCAGTGGCCCACTT 640503
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 640504 GCGGTGTCTACTTCTCCATGAAGACGCGTCTAGTCAAAATGGGCAATGGCAATATGTT 640563
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 640564 CCAGGATTATCTCTGGAGATGCGAGTTTACGCGCTTGTGCGCCAAACGCGAGTGGCGGA 640623
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 640624 ATTACGCGCTCTTCAATTAGAGGCTCAACACCTAGATTATCCCGCAATTCGTTAACTTA 640683
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 640684 ATGACATACCAAGTGGTTTCCAGGCAATCAAAAGTTATCCGTGTTCGAGATGACGTG 640743
QY 494 LeuGlnLysAlaLeuGluLeu 500
Db 640744 ATGAACAAATGTGAACCTTG 640764

RESULT 7

US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; TITLE OF INVENTION: Listeria innocua, genome, and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04

; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
US-10-398-221-2058
Alignment Scores:
Pred. No.: 1,098-35 Length: 3011208
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 16 Gaps: 11

US-10-009-823A-1 (1-502) x US-10-398-221-2058 (1-3011208)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 729803 ATGAATCAAACTATGTATACAGCTATTCTGGGATGAATGCGTTCACAAAGCATTATCA 729862
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 729863 GTGACATCAAAATAATATTGCCAATCGCAACACGACAGGCTATATAAAACAAGCGTCTGT 729922
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db 729923 TTCAATGATTACTTATCCAAACACAACAAATGGGATCGTTCGAGCGGACTTTATGCTGCA 729982
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 729983 ACACACCCCAATGAGCTTCGGTTCCGGTTTCGAAATTTGGGCAATTTTAAACCGATTATACA 730042
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
Db 730043 GCAGGTTCTCCACATCACTGGCAGAACAAAGATGCGACACTGCAAGGCGCGGCTTT 730102
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
Db 730103 TTCAATGCTGGCGATAACGCTGGTGGGAATATCGTTTACACACGACGAGCGTAGCTTGCA 730162
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgile 134
Db 730163 GTATCTGACATAACTATTATTAAACGGAACCTTGAACCAATTCAAATCCCAATTAATAGC 730222
QY 135 SerAsnAsnProAsnIleLysGlyLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 730223 GATAAAATGGCAACGCTTTAAACGGAACCTTGAACCAATTCAAATCCCAATTAATAGC 730282
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 730283 GCAATTCGCGGAGGACCAACAAAATGGTAGCTTAAGCGGTAACTTCCCACTTGATGG 730342
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSer 194
Db 730343 GCGGAAAAA---GATACGATTTCCTCCGAG----- 730369
QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 730370 -----CTT 730372
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
Db 730373 TCTGTATACGATAATGCTGGTGGAAAAACATAAACTTCAAGTCAATATGAAAGTGGACA 730432
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 730433 CCAGATGGGAGTGGCAATGTTTCTCTATGAATATGAATTCAAATG-----GAC 730480
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273

Db 5487 AATGCGCAGGGCGAATTAAACAAACCAACGACGCA----- 5519
 QY 294 AspLeuAsnAlaTprGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 Db 5520 ---CITAAAAATATCCAAATCAATTCACAGTAACGGC----- 5555
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 Db 5556 -----AAACAAGTCAAT 5567
 QY 334 MetTprAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 Db 5568 ATG-----GGCTTAACCTTAAGTGGCTTAACC 5594
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
 Db 5595 AACTAC-----GGAAACAACCA 5612
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 Db 5613 GTATTCTCACCACACTTCGACGGTAAGCGCTCGGACTGTAAGAAATATGCGAGTTACC 5672
 QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 Db 5673 GATTCTGGCTATATGCGAGTCAGTTACTCAGATGCTATATCCAGTTGCCCAACTT 5732
 QY 414 ProLeuAlaArgPheThrSerGlnAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
 Db 5733 GCGGTGGCTACTTCTCAATGAAGACGGCTTAGTCAAAATGGGGAACGCGAATATGTT 5792
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 Db 5793 CCAGGATTATCTTCGGCGTGCAGTATACGCGTTCGCCAAATGGCGTGGCGGA 5852
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 Db 5853 ATTACGGGTCTCTCTAGAGAGGTTCACACGTACACTTGTCCCGTGAATTCGTTAACTTA 5912
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
 Db 5913 ATGACATACCAAAATGGTTCCAGGGCATACAAAGTATTCTGTGTCGGCAACGCTG 5972
 QY 494 LeuGlnLysAlaLeuGluLeu 500
 Db 5973 ATGAATCAAACTGTGCACTTG 5993

RESULT 9

; Sequence 1, Application US/09790988
 ; Patent No. US20020127687A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHIGENOBU, SHUJI
 ; APPLICANT: WATANABE, HIDEMI
 ; APPLICANT: HATTORI, MASAHIRA
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
 ; FILE REFERENCE: 081356/0159
 ; CURRENT APPLICATION NUMBER: US/09/790,988
 ; PRIOR FILING DATE: 2001-02-23
 ; PRIOR APPLICATION NUMBER: JP2000-107160
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 640681
 ; TYPE: DNA
 ; ORGANISM: Buchnera sp.
 US-09-790-988-1

Alignment Scores:

Pred. No.: 2,4e-29 Length: 640681
 Score: 414.50 Matches: 132
 Percent Similarity: 39.96% Conservative: 73

Best Local Similarity: 25.73% Mismatches: 187
 Query Match: 16.03% Indels: 121
 DB: 9 Gaps: 13
 US-10-009-823A-1 (1-502) x US-09-790-988-1 (1-640681)
 QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 Db 372153 TCAAAATATATAGTATATAGTGGCTTACTAGCAAAATAATGATTACATGAAATATATACC 372212
 QY 24 AsnAsnIleAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAsp 43
 Db 372213 AATAATATTCTTAAGCATCAACTATAGATATAAATCTCGTAAACCTCTTTTTCAT 372272
 QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 Db 372273 ATGTTTCTCATTTTATTTCAAATACTACTAAT-----GGA 372311
 QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
 Db 372312 TACGAGTGGTATTTCAGTATATATACAAACCTTTAATAATAGGCATGTTAGTCAAACT 372371
 QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGln---ValThrLeuGlu 102
 Db 372372 GCACGAGATTTCGATTAGGAATTATAAAGACGGCTTTTCGGTCTTTGTAGACAGTCAA 372431
 QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122
 Db 372432 GGTCAATGCTATATATACAGAGATGGCAATTTCTTCCTCGATAAAGATCAAAATATTATC 372491
 QY 123 AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn 136
 Db 372492 AATATTCAAGGTATGATTACTAAGTCTTAATCATCTTCTCAAAAGTGAATTTAAT 372551
 QY 137 AsnProAsnIleLysGlyGluThrLeuGluProIleGlnLeuAspPheAsnAspProThr 156
 Db 372552 AAT-----AGATCCAACTTAGAACCTATTATTTA-----AAAAATTTCTAAT 372593
 QY 157 ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValAlaAsnLeuGly--- 175
 Db 372594 ATTTTAAAAAACAACTTCTCGAATCATGTATAAAGCGTTTGTGAATTCGTAATACT 372653
 QY 176 -----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe 189
 Db 372654 GAATCAAAAAGCAGTGTGGATAATTCTGACAACTATCTAAACCAAGAGATTACATG 372713
 QY 190 AlaLeuLeuGluSerTprLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSer 209
 Db 372714 ACTTATATTAGC----- 372725
 QY 210 TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr 229
 Db 372726 -----ATATATAATAAAGAGGAAAAAGAGATATTACTGTTCT 372767
 QY 230 PheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsn 249
 Db 372768 TTTAAT-----AAAAAGAAACAAATAAATGGACAGTAAT-----GTGGAATCAAT 372815
 QY 250 ProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly 269
 Db 372816 GATTCTGATATAAGAGACTATAAAAAAT-----AGTTTC 372851
 QY 270 ThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly 289
 Db 372852 GATTTAACGTTTAATGATGATGGCAATTAACCTCTCATATATGTTTTTAATTTACATCT 372911
 QY 290 SerAlaThrLysAspLeuAsnAlaTprGlnProAlaProLeuValAsnGlyLeuProGln 309
 Db 372912 AAAGATTCTAAAAAG----- 372926
 QY 310 PheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLys 329
 Db 372927 -----TATGAAATATACATTTTAAATTTA----- 372950

QY 330 SerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIle 349
Db 372950 ----- 372950
QY 350 GlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsn 369
Db 372951 -----ACAGGTACTATAGCAACAATCAAT 372974
QY 370 GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389
Db 372975 TCTGATGTTCTTGGGAAGAACATCTCAAAACGGATACCCCTCAAGGTAAATTTAAACACA 373034
QY 390 ValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAsp 409
Db 373035 TTGTATATCTTACTAATGCTGAATATTGGAACATATTGCAATCAAAAACACAAACA 373094
QY 410 PheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsn 429
Db 373095 ATAGGTCAAAATATTATCAAAATTTATCAATCCAGAAAAATACAACTGAAAGTGT 373154
QY 430 AsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu--ProGlyThr 448
Db 373155 AATTATGGTCTGCTACTGAGATCTAGGTGAAGCAAAACACCAATGAAGCGGTATT 373214
QY 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
Db 373215 CAAGAATCAGAGGTGTTAAGCAATAAAACGCTAGAGTATCAAAATGTTGATTTGAATAAA 373274
QY 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Db 373275 GAATTAATCAATATGATTAGCACACGTAATTTATCAATCTAACGCTCAATCTTTTAA 373334
QY 489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 373335 ACAGAATCAAAATAATTAATACATTAATAAATTACAG 373373

RESULT 10

US-10-335-977-56
Sequence 56, Application US/10335977
Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA

COMPUTER READABLE FORM:

ZIP: 02109-1875
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335.977
FILING DATE: 30-Dec-2002
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:
LENGTH: 666 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (5) LOCATION 1...666
SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-10-335-977-56

Alignment Scores:

Pred. No.:	2,99e-21	Length:	666
Score:	296.00	Matches:	78
Percent Similarity:	48.94%	Conservative:	37
Best Local Similarity:	33.19%	Mismatches:	98
Query Match:	11.45%	Indels:	22
DB:	16	Gaps:	3

US-10-009-823A-1 (1-502) x US-10-335-977-56 (1-666)

QY 272 ThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAla 291
Db 13 TCCTATAGAGTGAGGGGGCGGAATTTAGATGATCGTCCCTGAGCGCTGGGAA 72
QY 292 ThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSer 311
Db 73 TTAGTAGGGGGGTACGGCTAGGCTAATGTGTTGAAGGGGGCGGTTGCTATTTAAT 132
QY 312 AlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGln 331
Db 133 AACGACGGATCGCTTTCGGGCGATGAACCCGCTCTTTTGCATTTGACCCCTAAAAAT 189
QY 332 GlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLys 351
Db 190 -----GGCGCTGATGCC----- 201
QY 352 LeuProSerMetMetProIleGlnThrSerSerGlyAsnSer-----ThrAla 367
Db 202 ---CCCCAACGCATCAATTTAGCTTTTGGTTCTCAGGGAGCTTTGACGGATTGACAAGC 258
QY 368 ArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeu 387
Db 259 GTGGATAAGATTCTGAAACTTATCGGATTGAGCAAAACGGCTATCAACGGCGCGATTG 318
QY 388 ValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnVal 407
Db 319 ATGGATGTCGGCTTTGATCTGATGGGTGCTTTTAGGGGGCGTTTCAAGTAATGCGAGGACT 378
QY 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427
Db 379 TTAGCGCTCGCTCAAGTGGCTTTAGCGAATTTTCGTAACGATCGGGGCTTACAGCGTTTG 438
QY 428 GlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly 447
Db 439 GGAGCGAATGCTTTTCTCAAAACCGGAACTCTGGCAAGCCCTTAATCGGTGGCGCTAAT 498
QY 448 ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSer 467
Db 499 ACGGGCGGTAGGGGTTCAATTTCAAGATCCAACTAGAGTCTAGTAATGTGGATTGAGC 558
QY 468 ArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerVal 487
Db 559 CGAGGTTTAAAGAAATTTGATTGTGGTTCAAGGGGGGTTTCAAGCGAATCTTAAGCGGTA 618
QY 488 ThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 619 ACCACATCCGATCAAAATCTTAAACCCCTATTGAATCTCAAGCAA 663

RESULT 11

US-10-335-977-1
 ; Sequence 1, Application US/10335977
 ; Publication No. US20040052799A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; RELATING TO HELICOBACTER PYLORI FOR
 ; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
 CORRESPONDENCE ADDRESS:
 ADDRESS: LAHIVE & COCKFIELD
 STREET: 28 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: Windows NT 4.0
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/335,977
 FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/993,002
 FILING DATE: 17-Dec-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Mandragouras, Amy E.
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)742-4214
 INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 789 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Helicobacter pylori

FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...789
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-335-977-1

Alignment Scores:
 Pred No.: 2,618-20 Length: 789
 Score: 288.00 Matches: 107
 Percent Similarity: 28.68% Conservative: 39
 Best Local Similarity: 21.02% Mismatches: 109
 Query Match: 11.14% Indels: 254
 DB: 16 Gaps: 9

US-10-009-823A-1 (1-502) x US-10-335-977-1 (1-789)

QY 1 MetWetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 1 ATGCTCGCTCTCTATAGTGGCACTTCAGGATGCTGCCCAACAAACGACATTGAC 60
 QY 21 ThrValSerAsnAsnLeuAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 61 ACCACTCAACACACATCGCATGTCATACACCGGTTTAAATAATCTCGCGGAT 120
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 Db 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60

Db 121 TTTAACGACTTGTGTTTACCAAGCGATGCAATACGCGGCACCAACACGACGACT 180
 QY 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 Db 181 TTATCGCCAGATGGCATGGAAGTGGCTTAGCGGTACGCCCTAGTGCATACCAAAATG 240
 QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys 94
 Db 241 TTTTCGAAGCGACCCCTAAAGAAACGGAGATAATTTAGATATTGCTATTACAGGTAAA 300
 QY 95 GlyPheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
 Db 301 GGCTTTTCAAGTCAGCTTCTGATGGCACTACCGCTTACCAAGGAGCGGAAATTC 360
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 Db 361 AAGCTAGACGAGCAGGCAATCTTCTAAACAGCGAGGCTATCTCTCATC----- 411
 QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 Db 412 -----CCTCAATC-----ACTTTA----- 426
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 Db 427 -----CCCGAAGACACACGCAA-----GTGAAT 450
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
 Db 451 ATCGGT----- 456
 QY 194 SerTyrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 Db 456 ----- 456
 QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
 Db 457 -----GTGGATGGCAGC 468
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 Db 469 GTGACGCTGACT----- 480
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 Db 480 ----- 480
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 Db 480 ----- 480
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 Db 480 ----- 480
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 Db 480 ----- 480
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 Db 480 ----- 480
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
 Db 480 ----- 480
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 Db 480 ----- 480
 QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 Db 481 -----CAAGGCTTGCAAAACGACTTCTAACGTCGATCGGG-----CAAAATC 519

QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
 Db 520 ACTTTGGCTAATTTTGTCAATCCGGGGGCTTCAATTCATGCGGGGATAATTTGTTTCC 579
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 Db 580 ATCACAACGGTAGCGGGGATGCGATTGTGGGCAACCCGGATTCTCAAGCGTTAGGCAAG 639
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 Db 640 TTAGGCAAGGCTTTTTCGAGCTTAGTAACGTGAGATTGGTAGAAGAAATGACAGATCTA 699
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
 Db 700 ATCCCGCTCAAGGCGCTTAGAAGCAATTTCTAAAGCAATTTCAACCGCTGATGCCATG 759
 QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502
 Db 760 CTCCAACAGTCAATTCCTCTCAACCG 786

RESULT 12

US-10-335-977-2
 ; Sequence 2, Application US/10335977
 ; Publication No. US20040052799A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; RELATING TO HELICOBACTER PYLORI FOR
 ; DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 10031
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: Windows NT 4.0
 ; SOFTWARE: UNIX

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/335,977
 ; FILING DATE: 30-Dec-2002
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/993,002
 ; FILING DATE: 17-DEC-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mandragouras, Amy E.
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: GTN-018

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)742-4214

; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 816 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Helicobacter pylori

; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (B) LOCATION 1...816
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-335-977-2

Alignment Scores:

Pred. No.: 2,73e-20 Length: 816
 Score: 288.00 Matches: 107
 Percent Similarity: 28.68% Conservative: 39
 Best Local Similarity: 21.02% Mismatches: 109
 Query Match: 11.14% Gaps: 254
 DB: 16

US-10-009-823A-1 (1-502) x US-10-335-977-2 (1-816)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
Db	28	ATGCTCCGCTCTCTATAGTGCACCTTCAAGGATGCTCGCCCAACAAACGACATGAC	87
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnVal	40
Db	88	ACCACTTCAAAACATCGCCATGTCAATACACCGGGTTTAAAAAATCTCGCGGGAT	147
QY	41	PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn	60
Db	148	TTTAAACGACTTGTTTACCAAGCGATGCAATACCGCGCACCAACACACACACGACT	207
QY	61	-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle	74
Db	208	TTATCGCCAGATGCGCATGGAAGTGGCGCTTGGCGTAGCCCTAGTCGATTACCAAAATG	267
QY	75	PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys	94
Db	268	TTTTTCGACGCGCCCTAAAGAAACGAGAAATAATTAGATATTCTATTACAGGTAA	327
QY	95	GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsn	113
Db	328	GGCTTTTTCAGTCCAGCTTCTGATGGCATTACCGCTTACACAGGAGCGGAATTC	387
QY	114	ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySer	133
Db	388	AAGCTAGCAGCGAGCGCAATCTTGTAACAGCGGGCTATCTCTCATC-----	438
QY	134	IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAsp	153
Db	439	-----CCTCAATC-----ACTTTA-----	453
QY	154	AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaVal	173
Db	454	-----CCGAGACACACCGCAA-----GTGAAT	477
QY	174	LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeu	193
Db	478	ATCGGT-----	483
QY	194	SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAla	213
Db	483	-----	483
QY	214	MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAsp	233
Db	484	-----GTGGATGGCAGC	495
QY	234	ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSer	253
Db	496	GTGAGCGTGACT-----	507
QY	254	GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMet	273
Db	507	-----	507
QY	274	SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAla	293
Db	507	-----	507
QY	294	AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSer	313
Db	507	-----	507

QY 314 PheValGlyAlaGlyLeuGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 507 -----
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 507 -----
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
Db 507 -----
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValThrIleThr 393
Db 507 -----
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 508 -----CAAGGCTTGCAACGACTTCTACGTCATCGGG-----CAAAATC 546
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 547 ACTTGGCTAATTTGCTCAATCCGGCGGGCTTCACTATGCGGGATAATTTGTTTTC 606
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 607 ATCAACCAACGCTAGCGCGCATGCGATTGTGGCAACCGGATTCTCAAGGCTTAGGCAAG 666
QY 454 LeuSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 667 TTAGGCAAGCTTTTGGAGCTTAGTAACGTAGATGTGTAGAGAAATGACAGATCTA 726
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 727 ATCACCCTCAAGGGCTTAGTAAGCAATTTCTAAAGCATTCATAACCGCTGATGCCATG 786
QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502
Db 787 CTCCAACAGTCAATTCCTCAACGC 813

RESULT 13

US-10-335-977-3

; Sequence 3, Application US/10335977

; Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
APPLICATION DATA:
FILING DATE: 08/993,002

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 837 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...837

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-10-335-977-3

Alignment Scores:

Pred. No.: 2,848-20 Length: 837

Score: 288.00 Matches: 107

Percent Similarity: 28.68% Conservative: 39

Best Local Similarity: 21.02% Mismatches: 109

Query Match: 11.14% Indels: 254

DB: 16 Gaps: 9

US-10-009-823A-1 (1-502) x US-10-335-977-3 (1-837)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 52 ATGCTCCGCTCTCTCTATAGTGCACCTTCAGGATGCTCGCCCAACAAACGACATTGAC 111
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 112 ACCACTTCAACAAACATCGCAATGTCAATACCAACCGGGTTTAAAAATCTCGCGCGAT 171
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 172 TTTAACGACTTGTTTTACCAAGCGATGCAATACCGCGCACCAACACAGCAACACGACT 231
QY 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
Db 232 TTATCGCCAGATGCGATGGAAGTGGCGCTTGGCGCTAGTGGCTAGTGGATTTACCAATG 291
QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
Db 292 TTTTCGCAAGCGACCCCTTAAGAACCGAGAAATATTAGATATTGTTATACAGTAAA 351
QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
Db 352 GGCCTTTTTCAGTCCAGCTTCTCTGATGGCACTACCGCTTACACAGGAGCGGGAATTC 411
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 412 AAGTAGACGACGAGGCAATCTTGTAAACGAGGAGGCTATCTCTCTATC----- 462
QY 134 IleSerAsnAsnProAsnIleLysLysGluThrIleGluProIleGlnLeuAspPheAsn 153
Db 463 -----CCTCAATC-----ACTTTA----- 477
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 478 -----CCGGAAGACACACCGCAA-----GTGAAT 501
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
Db 502 ATCGGT----- 507
QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
Db 507 ----- 507

QY	214	MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla	233
DB	508	-----GTGGATGGCAGC	519
QY	234	ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp	253
DB	520	GTGAGCGTGACT	531
QY	254	GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe	273
DB	531	-----	531
QY	274	SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys	293
DB	531	-----	531
QY	294	AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn	313
DB	531	-----	531
QY	314	PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyLeuLysSerGlnGlnAsn	333
DB	531	-----	531
QY	334	MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro	353
DB	531	-----	531
QY	354	SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer	373
DB	531	-----	531
QY	374	ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr	393
DB	531	-----	531
QY	394	SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle	413
DB	532	-----CAAGCCTTGCAACGACTTCTAACGTCGCGG	570
QY	414	ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer	433
DB	571	ACTTTGGCTAATTTTGTCATCCGCGGGCTTCATTATGGGGGATAATTGTGTTTCC	630
QY	434	AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys	453
DB	631	ATCACCAACGCTAGCGCGCATCGGATTGGGCAACCGGATCTCAAGGCTTAGGCAAG	690
QY	454	LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet	473
DB	691	TTAAGGCAAGCGCTTTTGGAGCTTAGTAACTGTAGATTGGTAGAAGAATAACAGACTCA	750
QY	474	IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet	493
DB	751	ATCACCGCTCAAGGGCTTATGAAGCCAAATTTCAAAGCATTCAAAACGCTGATGCCATG	810
QY	494	LeuGlnLysAlaLeuGluLeuLysArg	502
DB	811	CTCCAAACAGTCAATTCCTCCATCAACGC	837
RESULT 14			
US-09-974-300-1888			
; Sequence 1888, Application US/09974300			
; Patent No. US20020146721A1			
; GENERAL INFORMATION:			
; APPLICANT: Berk, Randy M.			
; APPLICANT: Clausen, Ib Groth			
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene			
; TITLE OF INVENTION: Expression			
; FILE REFERENCE: 10085.500-US			
; CURRENT APPLICATION NUMBER: US/09/974,300			
; CURRENT FILING DATE: 2001-10-05			

;; PRIOR FILING DATE: 2001-03-27

;; NUMBER OF SEQ ID NOS: 8481

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 1846

;; LENGTH: 1524

;; TYPE: DNA

;; ORGANISM: Bacillus licheniformis

US-09-974-300-1846

Alignment Scores:

Pred. No.: 2,46e-12 Length: 1524

Score: 215.50 Matches: 130

Percent Similarity: 37.21% Conservative: 81

Best Local Similarity: 22.93% Mismatches: 207

Query Match: 9.33% Indels: 149

DB: 28 Gaps: 28

US-10-009-823A-1 (1-502) x US-09-974-300-1846 (1-1524)

```

QY 5 LeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsn 24
DB 25 CTGGAATTCGCAAAAGAGGTTGTCCGCGAGCAATCAGCTCTGAGCGTGACAGCAAC 84
QY 25 AsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGln----- 42
DB 85 AATATATCCACGCCAATACCAAGAGGTATTCAAGACAGCGGTGTATCATTTAAGCATCA 144
QY 43 -----AspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyPro 59
DB 145 ACCCGTTATCCGTCGTCATCAAGAGAC---TCTGTCGACTTGCGAGC----- 189
QY 60 AsnGlnAlaGlyMetGlyAlaGlnValGlySerValargThrIlePheThrGlnGlyAla 79
DB 190 ---CAGATGGGTACAGCGGTGAATCGTTGCGTTGAAAGGGTT-----AGCGACAGC 240
QY 80 PheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnVal 99
DB 241 TTT-----LeuAsnAsp-----CTG 246
QY 100 ThrLeuGluAspLysValHisTyrArgAlaGlyAsnPheargPheThrGlnAspGly 119
DB 247 GACTATCAATACAGCGCAGATACAAAGCGCGCTATTACAAATGCAAAAGTGGACGCT 306
QY 120 Phe-----LeuAsnAsp-----123
DB 307 TTTAACCAAAATGAAGGATTATGAATGAATGAACGACACCGGACTAAACAAAGTTCTG 366
QY 124 -----ProSerGlyPhe 127
DB 367 AATTCATCTGGAATTCATTCAGGAATTCAGCAACAAACCCGACGAGCGCTGCT 426
QY 128 ThrLeuMet-----GlySerArgIleSerAsnAsnProAsnIleLysLysGluThr 144
DB 427 TCTCTTGTGCGCAAAAGGCGCGGTGCGCGATACATTTAACACTCTGTATGATCG 486
QY 145 LeuGluProIle-----GlnLeuAspPheAsnAspProThrValAla 158
DB 487 CTGCAACACCGGTCAATCAAAATTTAGCGCACAGATTGATCAAAATGTGATGACGATT--- 543
QY 159 LysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal-----172
DB 544 AACTCGTTTTTCCCGCTGACAGCTGAACAGCAAAATCCGCAAGTCGACGCGCAAC 603
QY 173 -----AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
DB 604 GGCTATCTCCGCAACGATTTGTACGATAAGCGGTGATCAGCTCTTGTATGAACTGTCTCA 663
QY 188 TyrPheAlaLeuGluSerThrLysGlyAsnGlyThrProProIleSerThrSerAsn 207
DB 664 ATGGTAAATCAAGTCAGTCAGTAAGCAACCGCGGCAATCCACTCCGCAACCGCTGAA 723
QY 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227

```

```

DB 724 GGCATTGTGTCA-----ATAGAAGTCTCTTGAGTAAACGGCCAGTCT-----CTGGGC 771
QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
DB 772 AAGTTCTTGACGGGCCAAACATTCACGACTGAACCT-----GTCAAA 813
QY 248 MetAsnProSerGluAsp-----GlySerAlaAlaSerGlyThrAsp---261
DB 814 GTGAATTACGACACAGATACAGCCCTTGACGGGGTTTCCCTCGCGGGAGCGACATC 873
QY 262 -----SerAlaGlyLeuLeuMetSer-----GlyThrMetThrPhe 273
DB 874 GGCATCGATTCTGTTTACTAGCAAAAGGTCCTCTCGGTTTAAATCGAATCTTACGGATAT 933
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 934 ATGTCAAACGGTGAAGAAAGAGTTG-----TATCCGAAATGCTCCGATCTG 984
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 985 GATACGATCGCGCTTCTTTTCGAAAGAGATTAAATGCGATCCATCAA-----AGCGGG 1038
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
DB 1039 TTTACTTCATCGGGA---GAACCGGGAGGCAATTTCTTTTGATTTTACCGGAGGAG---1092
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 1093 -----GCTGAGCTTCAAGAGGAGCGCC-----GGCAAAATCAAA 1128
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg-----AsnGlySer 371
DB 1129 GTAGCCGATGATCATGAATTCAAAGGCGATAAAGTCGCGCATCGCTGAATGGTGAA 1188
QY 372 SerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThr 391
DB 1189 GCGAGC-----GACAAACGCCAAGCGACAAATCTCGCTAATGCTCTC 1230
QY 392 IleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyr 411
DB 1231 ACCAAG-----AAATTCNAATCGGTGACAAAACGACGACCGCTTTAGACTATTAC 1281
QY 412 AsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlyGluGlyAsnAsnHis 431
DB 1282 -----GCAGGCGATCATCGTGAATGGCGTAAAGCTCAGGAACAAACAGA 1329
QY 432 TyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyr 451
DB 1330 CTGCGCAAAATACAGAAACG-----CTTGTAACACACTGCTGAATTG 1371
QY 452 GlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetVal 471
DB 1372 AACGACAGCTCTGT-----AGCGCGGTATCATTTAGCAGAAATATGTCG 1416
QY 472 AsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAsp 491
DB 1417 AACATGATCCAAATTCAGCATCTTATAACGGCGCTCGGAGAAATGATTACGCTGCAGGAC 1476
QY 492 ThrMetLeuGlnLysAlaLeu 498
DB 1477 GAAGTCTTCGATAAGATTATT 1497

```

RESULT 16

US-10-335-977-78

; Sequence 78, Application US/10335977

; Publication No. US20040052799A1

; GENERAL INFORMATION:

; APPLICANT: DOUGLAS SMITH et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

; RELATING TO HELICOBACTER PYLORI FOR

; DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 10031

; CORRESPONDENCE ADDRESS:

349 GGGNACGCCCTAACCAATCTCATTTTAAATGCAAGAGGCTATTCTATGAACGCTAGAGCC 408
487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuLysArg 502
409 TTTGGCGGCGGCGATGATCATGATCAAGAGGCTATTAGCTTGAAGAAA 456
RESULT 17
US-10-210-296-2
; Sequence 2, Application US/10210296
; Publication No. US20030021802A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED
; TITLE OF INVENTION: METHODS
; TITLE OF INVENTION: AND MATERIALS
; FILE REFERENCE: PCI0589A
; CURRENT APPLICATION NUMBER: US/10/210,296
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/689,065
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 5445
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-10-210-296-2
Alignment Scores: 3.39e-09 Length: 5445
Pred. No.: 193.00 Matches: 40
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.46% Indels: 0
DB: 14 Gaps: 0
US-10-009-823A-1 (1-502) x US-10-210-296-2 (1-5445)
QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482
DB 3 AACGTAGACATGAGCAGAGAAATGGTTAATATGATTATTACACGGTGGTTTCAGAGT 62
QY 483 AsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuLysArg 502
DB 63 AATAGTAATCTGTTACACAGACACACATGCTACAAAGACACTTGAACCTAAAGCGT 122
RESULT 18
US-10-449-462-2
; Sequence 2, Application US/10449462
; Publication No. US20030202983A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products, Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS AND RELATED METHODS AND
; TITLE OF INVENTION: MATERIALS
; FILE REFERENCE: 3153.00187/PCI0589A
; CURRENT APPLICATION NUMBER: US/10/449,462
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US/09/689,065
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US Prov. 60/160,922
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US Prov. 60/163,858
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 5445
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-10-449-462-2
Alignment Scores: 3.39e-09 Length: 5445
Pred. No.: 193.00 Matches: 40
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.46% Indels: 0
DB: 14 Gaps: 0

ADDRESSER: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 459 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...459
SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-335-977-78
Alignment Scores: 1.27e-11 Length: 459
Pred. No.: 201.50 Matches: 47
Score: 53.68% Conservative: 26
Best Local Similarity: 34.56% Mismatches: 52
Query Match: 7.79% Indels: 11
DB: 16 Gaps: 3
US-10-009-823A-1 (1-502) x US-10-335-977-78 (1-459)
QY 378 SerGlnAspGlyTyrProGlnGlyAspLeuValThrThrSerGluGlyLys 397
DB 49 AGTCAGACGCAAGCCTAGGGGCAATTTTAGACACATGCGCATTTGAAGAAATGGCGTG 108
QY 398 LeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArg 417
DB 109 ATTTCTAGCCTTTAGTAACGGGCTGTAGAGCGGTCGCTCGCATCGGTATTTAGCT 168
QY 418 PheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyr-----SerAla 434
DB 169 TTCACACGATCAAGGCTTAAAGGAATCGCGGTAACTCTATGAATGCAAGAGGC 228
QY 435 ThrLeuAsp-----SerGlyGlyProGluPheGlyLeu-----Pro 446
DB 229 ACCATTATGCGAAACACAGACCCCTAAGCGTAACCCCATTTTAGGCTGGGACGAAGAG 288
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnIleuGluThrSerAsnValAspMet 466
DB 289 GCGACAGCTCAAGTTGGGAAATCAGGCACAAATATTAGAAACGACGACGTAATGCC 348
QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486

Score: 193.00 Matches: 40
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.46% Indels: 0
 DB: 15 Gaps: 0

US-10-009-823A-1 (1-502) x US-10-449-462-2 (1-5445)

QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482
 Db 3 ACGTAGACATGAGCAGAGAATGTTAATGATATTATTCACGCGGTTTTCAGATG 62
 QY 483 AsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
 Db 63 AATAGTAATCTGTACACAGCAGACACAACTGTACAAAAGACACTTGAACTAAGCGT 122

RESULT 19

US-10-335-977-82
 ; Sequence 82, Application US/10335977
 ; Publication No. US20040052799A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; RELATING TO HELICOBACTER PYLORI FOR
 ; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 28 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109-1875

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: Windows NT 4.0
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/335,977
 FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: GTN-018
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 1395 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...1395

SEQUENCE DESCRIPTION: SEQ ID NO: 82:

US-10-335-977-82

Alignment Scores:

Pred. No.: 1.12e-09 Length: 1395
 Score: 189.50 Matches: 59
 Percent Similarity: 43.14% Conservative: 29
 Best Local Similarity: 28.92% Mismatches: 81

Query Match:

DB: 7.33% Indels: 35

16 Gaps: 6

US-10-009-823A-1 (1-502) x US-10-335-977-82 (1-1395)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHiserThrGlyLeuGly 20

Db 1 ATGAACACACACCTTATTAACCGCTTATAGCGGATCAAGACCCACAGCTTGGTTATGAC 60

QY 21 ThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40

Db 61 AGCCTTCCAAATAATATCGCAATGCTCAATACTTTAGGCTATCGCTCAATATCGGAG 120

QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60

Db 121 TTAAACCTTGTTCCTTCGCAATTTAGACGGCTTTGAACGCCCAATCCGTGTGGCTAAC 180

QY 61 GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGln---GlyAla 79

Db 181 GACCGAAATTTACGGCGTT---ACAGGATCAGGCAATGCTTCTTAATAAAGACGGCGAA 237

QY 80 PheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnVal 99

Db 238 TACATGCCAAGTGAAGGGGAATTCACATGCGGTATCAGGGCAAGGTTGTTGTGATA 297

QY 100 -----ThrLeuGluAspLys 104

Db 298 GGGCCCAATAAAACGGGGAATGACCAATTAATAAGATGCTTTAGCAAAAACAGGAT 357

QY 105 ValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspPro 124

Db 358 AATTTCCTTAACCGGTGCGGCAATTTCCGCGGAGACGCGGATGCTATTATTAACCCCT 417

QY 125 SerGlyPheThrLeuMetGly-----SerArg 133

Db 418 GAGGCTATATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 477

QY 134 IleSerAsnAsnProAsnIleLysLys-----GluThrLeuGluProIleGlnLeu 150

Db 478 ACCGTAGATAGATGAACATTTGCAAAAATTCGATAAACACCCCTTTCGCCCTTACAATC 537

QY 151 AspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170

Db 538 -----CCCCAGATTGACTTACAGCGCGGTGCTTAGCAGAAAGTGAATATC 585

QY 171 ValValAsnLeu 174

Db 586 AGCGTGAATCTA 597

RESULT 20

US-10-369-493-28400

; Sequence 28400, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 28400

; LENGTH: 1965

; TYPE: DNA

; ORGANISM: Burkholderia fungorum

US-10-369-493-28400

Alignment Scores:

Pred. No.: 159e-08 Length: 1965
Score: 180.50 Matches: 144
Percent Similarity: 34.25% Conservative: 93
Best Local Similarity: 20.81% Mismatches: 224
Query Match: 6.98% Indels: 233
DB: 15 Gaps: 28

US-10-009-823A-1 (1-502) x US-10-369-493-28400 (1-1965)

QY 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 1 ATGTCGAATCTCATCAATCTCGGCCTCAGTGGACTGAACGCGACCCAGTGGGACTCAG 60
QY 21 ThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 61 ACCACGGCCAGAAATCAGAACGGTTCGACCGCGGTACACCATCGAAACCGCGTA 120
QY 41 PheGlnAsp 43
Db 121 TATCGGAAGCGCGGCCAGTACACGGTTCGGGTATCTGCGCAAGGGGTCTCGACC 180
QY 44 -----LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGly 58
Db 181 ACACCGTGTACGCCAGTACAGCCAGTATCTACCACTGAGCTGAACACGGCGAGTCG 240
QY 59 ProAsnGlnAla 62
Db 241 TCACGAGTTCGTGTGCGAGTACACACGCTGATCTCGCAACTGAACATCTGATCGGC 300
QY 63 GlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluPro 82
Db 301 AGCCGCACTCCGGGATTCGCGCGGATCACCAGCTATTTCACCGCTTGCAACACGTC 360
QY 83 GlyAsnSerValThrAspLeuAlaIle-----GlyGlyLysGlyPhe 96
Db 361 TCGAACACGCTCGACCTCGGACGCGGCACACCGCGATGAGCGGCGCAG----- 414
QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheTh 116
Db 415 -----ACGTGTGTGAACCATCAACGCGCGGTTCAGCAATACGACGCGTCCGC 465
QY 116 rGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAs 136
Db 466 CAGACGCTCAACACGCAACTCACCACACCGTTTCGACGATTACAGCTACTCGCAGCAG 525
QY 136 nAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProTh 156
Db 526 ATCGCGCACTGACGACGACAGATCGCCCGCAGCGCAGCAGC----- 565
QY 156 rValAlaLysSerProAlaLysThrSer-----ThrAlaLeuAsnAlaValValas 173
Db 566 -----AAGGCGACGCGCGCAACCATGTGTCGACCGCGCATCTC-GCCGTGTCCAA 617
QY 173 nLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
Db 618 CTTGTCGAGTTGATCGCGTGAACGTCGTCGACGACGACGACGACGACGACGACGAC 677
QY 193 uSerTyrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPr 213
Db 678 GAGC-----AATGGCCACGCGCTCGTCTGCTCCACCAAC-----AGCTAC-----AA 719
QY 213 oMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAl 233
Db 720 CTTGGGCAACCGGCTTCCACCGGCGCACAGTGAATGTCTGTCGATCTCTCGGCA 779
QY 233 aProSerThrGlySerIysThrPheGluTyrLeu----- 245
Db 780 GCGCGCGCAACCG 839
QY 245 ----- 245
Db 840 CACGCTCGGCGCTGCTTGTGCTTCGCGACGCGACGCTCGATCCGCGGCGAGCGCACT 899

QY 246 -----ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly----- 259
Db 900 CGGCGCGATCGCCGTGAGCTTCTCCGCGCAGGTCAACGCGCAGACGGGCTCGGTATCAC 959
QY 259 ----- 259
Db 960 GTCGCGCGGCCCAAGCGCGCGCTGTTCTCGGTGGCGCGCCGACCGTCTTACGGCAA 1019
QY 260 -----ThrAspSerAlaGlyLeuLeuMetSer----- 268
Db 1020 CACACAGAACCCCGCACCGGTCTGTAAGCTGTCTGTCGCGGACGCCACGACGCCGAC 1079
QY 269 -----GlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThr-----Al 284
Db 1080 CACCGCGACTACACGCTGGCTACAAACGCGCACCACTACACGCTGACCGCAATTCGAC 1139
QY 284 aPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVa 304
Db 1140 GGGCACCGTGTGGTTCGGCACT-----AACCTGAGCCAGCG-----AT 1181
QY 304 lAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLe 324
Db 1182 CAACGGCTG-----AATTTCTGACCAACCGCG-----ACGAT 1214
QY 324 uAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaA 344
Db 1215 GAACGCGCGGATTCGTTTCAGTCGACGCGCAGC-CGCGCGCGCTGAACTGTTCCGCA 1273
QY 344 lalleGlyThrAspIleGlyLysLeuProSerMetMetProIle-----G 359
Db 1274 CGCCACACCGACGATCGCGCTGCGCGCGCGCGCGCGCGCTGCTGCGCGCGCGCG 1333
QY 359 lNThrSerSerGlyAsnSerThrAlaArgAsnGlySer----- 371
Db 1334 CGAGCAACACCGCACCGCCACGATTACGAGGCGCAGCGTCAACGCGCGCTACACCATGC 1393
QY 372 --SerSerThrArgArgTyrSerGlnAsp-----GlyTyrProGlnGlyA 386
Db 1394 CGAATCTCACCAACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1453
QY 386 spLeuValAspVal----- 390
Db 1454 CGACGCTGACGCTGCGCGCTACGCTCCACCAACGTCACGATCCGACGCGCGCGCG 1513
QY 391 -----ThrIleThrSerGluGlyLysLeuGlnGlyL 401
Db 1514 TGGTCCGCTATTCGTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1573
QY 401 ySyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla-----ArgPheT 419
Db 1574 AGATGACAAATGTGTCGCTGACG--ATCAGCAACGCGCACCGCTCCCGCGCGCAAGTTCA 1630
QY 419 hr-----SerGluAspGlyLeuArgGluGlyAsnAsn-----HisT 432
Db 1631 CCATCGCGCGCAACACCGCGCGCACCAACGCGGTGCGCAACGCGCTGCGCATTTGTCGAAT 1690
QY 432 yzSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly----- 447
Db 1691 TGTCCACCGCAAGCGATGTGCGCGGAGCGGTCACTGACGCGCGCGGTATGCGAACT 1750
QY 448 --ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn----- 463
Db 1751 AGTCAACAATATCGCAAC---CAGACCAACCATCCAGACCTCGAGCGCGCGCAGA 1807
QY 464 -----ValAspMetSerA 468
Db 1808 GTTTCGCTGTCGACGATCACCACCGCGCAGCTCGTTCGCGCGGTGAACATCAACG 1867
QY 468 rGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValT 488
Db 1868 AAGACGACCAACCTGCTTTCAGTATCAGCAACTGTATCAGCGCAACACGAAAGTCACTCC 1927

Qy 391 -----ThrIleThrSerGluGlyLysLeuGlnGlyL 401
Db 1514 TGTGCGGTATTCGTGGCGACCGCGCCACGCTCAGATCAACAACGCGCGCGCC 1573
Qy 401 ystYrSerAsnSerGlnValValAspPheTyraenileProleuAla-----ArgPheT 419
Db 1574 AGATGAACAATGTGTGGTGACG---ATCAGCAACGACCGGCTGCGCGCAAGTTCA 1630
Qy 419 hr-----SerGluAspGlyLeuArgGluGlyAsnAsn-----HisT 432
Db 1631 CCATGGCCCGAACAACCGCGCGACCAACGCGCTGCGAACGGCTGGCATTTGCGAAT 1690
Qy 432 yrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly----- 447
Db 1691 TGTCCACCGCGAAAGCGATGTCGGCGGAGCGGTCAAGCTGACGCGCGGTATGCGAAT 1750
Qy 448 --ThrSerAsnTyrrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn----- 463
Db 1751 ACGTCAACAATATCGGCAAC---CAGACCAACAGATCCAGACCTCGAGCGCGCGCAG 1807
Qy 464 -----ValAspMetSerA 468
Db 1808 GTTCGTGTGACGACATCAACCGCGCGAGCGTGGTTTCGGCGGTGAACATCAACG 1867
Qy 468 rGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValT 488
Db 1868 AAGAAGCAGCGCAACCTGCTTCAGTATCAGCAACTGTATCAGCGGAACAGCAAGGTCTATCC 1927
Qy 488 hrThrAlaAspThrMetLeuGlnLysAlaLeu 498
Db 1928 AGACCGCGCAGACCTGTTCACACGATCTC 1959

RESULT 22

US-10-193-764-64
; Sequence 64, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 3285
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-193-764-64

Alignment Scores:
Pred No.: 1,78e-07 Length: 3285
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 15 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-193-764-64 (1-3285)

Qy 30 AsnThrIleGlyTyrrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 1807 AATATTTTCAGGTTTCAATATGACGAGATTAACACTAAAGAT-----GGTAGTGATTTA 1860
Qy 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
Db 1861 ACTATTGGTACACCAATAGTGTGATGGTACTAAT----- 1896

Qy 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
Db 1897 ---GCCAAAAAGTAACCTTT-----AACCAAGTTAAAGAT 1929
Qy 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrrThr 108
Db 1930 TCAAAAATCTCTGTGACGGT---CACAAGGTGACACTACACGCAAGTGGAA----- 1980
Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 1981 ACATCCCGTAGTAATAACAACACTGAAGATGACAGTACATAAT---GCCGGCTTAAT 2037
Qy 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
Db 2038 ATCGATGCAAAAAATGTAACACAGTAAACAATAATTAATCTCTCAAAAGCAGTACAGCATC 2097
Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 2098 -----TCTGCCACAGTGGAGAAATACCCTAAACAGGTACACCAT 2142
Qy 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrr 188
Db 2143 AACGCAACCACTGGTAACTGGAGATAACCGCTCAACAGGTAGT-----ATC 2190
Qy 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr-----ProIle 203
Db 2191 CTAGTGGAAATTTAGTCCAGCTCTGGCTCTGTCTTAACACTTACTGCAACCGGCGCTCTT 2250
Qy 204 SerThrSerAsnTyrrSerTyrrAlaGlnProMetArgValTyrrAspGlnGlnGlyAsnSer 223
Db 2251 GGTGTAAAGCAATATTTCC-----GGCAACACC 2277
Qy 224 HisAlaIleThrValTyrrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 2278 GTTACTGTTACTGCAAAAT---AGCGGTGCATTAACCACTTTGGCAGCGGTCTCAATTA 2334
Qy 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 2335 GGAACCGAGAGTGTAAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTCTCGT 2394
Qy 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
Db 2395 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAATTCAAAATTTAA 2454
Qy 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300
Db 2455 GCAACAACAGCGCGAGGTAAAGTAAACAGTGCACAGGTACAAATTTGGTGGTACGATTTC 2514
Qy 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
Db 2515 GGTAATACGGTAAAT-----GTTACGGCAACCGCTCGC 2547
Qy 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla 340
Db 2548 GATTTAAACAGTTGGGAATGGCGCAGAA-----ATTAAATGCGACAGNA 2589
Qy 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db 2590 GGAGCTGCAACCTTAACCTACATCATCGGCAAAATTA-----ACT 2628
Qy 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgTyrrSerGlnAsp 380
Db 2629 ACCGAAGCTAGTTTCACACATTACTTACGCCAAGGCTAGGTAATCTTTTCAGCTCAGAT 2688
Qy 381 GlyTyrrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 2689 GGTAGCGTTGACGAGGATTAATGCGCCCAATGTGACACTAAATTAACAGGCACTTTA 2748
Qy 399 -----GlnGlyLys-----TyrrSerAsnSerGln 406
Db 2749 ACTACCGTGAAGGTTCAAACTTAATGCAACCGCGGTACCTGCTGTTATTAAACCAAAA 2808
Qy 407 ValValAspPheTyrrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426

QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
 Db 4629 TCAAAAACGGTATAAACACCGCTACTGTTA---AAAGCGTTTAAATTTGATGTGAATAC 4685
 QY 487 ValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
 Db 4686 ATTCAACCGGTATAGCAAGCGTAGATGAAGTAATTGAACGCAACGC 4733

RESULT 24

US-10-193-764-62
 ; Sequence 62, Application US/10193764
 ; Publication No. US20030133943A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomis, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
 ; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
 ; FILE REFERENCE: 1038-1239MIS
 ; CURRENT APPLICATION NUMBER: US/10/193,764
 ; CURRENT FILING DATE: 2002-07-12
 ; PRIOR FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 91
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 62
 ; LENGTH: 5116
 ; TYPE: DNA
 ; ORGANISM: Haemophilus influenzae
 US-10-193-764-62

Alignment Scores:
 Pred. No.: 3,36e-07 Length: 5116
 Score: 173.50 Matches: 110
 Percent Similarity: 37.30% Conservative: 75
 Best Local Similarity: 22.18% Mismatches: 210
 Query Match: 6.71% Indels: 101
 Ds: 15 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-193-764-62 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
 Db 3480 AATATTTTCAGTTTCAATAAAGCAGATACAGTAAAGAT-----GGTACTGATTTA 3533
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
 Db 3534 ACTATTGGTAACCAATAGTCTGCTGCTACTAAT----- 3569
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
 Db 3570 ---GCCAAAAGTAACCTTT-----AACCGGTTAAAGAT 3602
 QY 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
 Db 3603 TCAAAAATCTCTGCTGACGGT---CACAAAGTGACACTACACAGCAAGTGGAA----- 3653
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 Db 3654 ACATCCGGTAGTATAACACACTGAAGTAGAGTACACATAAT---GCCGGCTTAAC 3710
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
 Db 3711 ATCGATGCAAAAATGTAACAGTAACAAACAAATATTACTTCTCAAAAAGCAGTGGAGCATC 3770
 QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
 Db 3771 -----TCTGCCACAGTGGGAGAAATTACCACTAAACAGGTACACCACT 3815
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 Db 3816 AACGCAACCACTGGTAACGTGGAGTAAACCGCTCAACACAGTACT-----ATC 3863

QY 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr-----ProProlle 203
 Db 3864 CTAGGTGGAATTTGATCCAGCTCTGGCTCTGCTAACAATCTCTGCAACCGAGGGCTCTTT 3923
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
 Db 3924 GCTGTAGCAATATTTCG-----GGCAACAC 3950
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
 Db 3951 GTTACTGTACTGCAAAAT---AGCGGTGCATTAACCACTTTGGCAGGCTCTACAAATAA 4007
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
 Db 4008 GGAACCGAGAGTGTACCACTTCAAGTCATCAGCGCATATCGCGGTACGATTCTGGT 4067
 QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGlyLeuLys 280
 Db 4068 GGCACAGTAGAGGTAAAGCAACCGAAAGTTTAAACCACTCAATCCAAATTCAAAATAA 4127
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300
 Db 4128 GCAACACAGCGGAGGTAAAGTAAAGTCAACAGTCAACAGTCAACATTTGGTGTACGATTTC 4187
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 Db 4188 GGTAAATACGGTAAT-----GTTACGGCAACCGCTGGC 4220
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla 340
 Db 4221 GATTTAACAGTTGGGAATGGCGCAGAA-----ATTAATGCCACAGAA 4262
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 Db 4263 GGAGCTGCAACTTAACCTACATCATCGGCAAAATTA-----ACT 4301
 QY 361 SerSerGlyAsnSerThrAlaAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
 Db 4302 ACCGAAGCTAGTTTCAACATCTTACGCCAAGCGTACAGTAAATCTTTTCAGCTCAGGAT 4361
 QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGlyLysLeu 398
 Db 4362 GGTAGCGTTGCGAGGAGTATTAAATGCCCAATGTGACACTAAATACTACAGGCACTTTA 4421
 QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
 Db 4422 ACTACCGTGAAGGTTTCAACATTAATGCAACACCGCGTACCTGTGTTTATTAAACGCAAAA 4481
 QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
 Db 4482 GACGTGAGCTAAATGGCGCAGCATTTGGGTACCAACAGTGTAAATTCACACCAACGCA 4541
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
 Db 4542 AATGGCTCCGCGCAGCGTAAATCGCGCAACCTCAAGC-----AGAGTGAACATCACT 4592
 QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
 Db 4593 GGG-----GATTATCAATAATTAATGATTA-----AATATC---ATT 4628
 QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
 Db 4629 TCAAAAACCGGTATAACACCGCTACTGTTA---AAAGCGTTTAAATTTGATGTGAATAC 4685
 QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
 Db 4686 ATTCAACCGGTATAGCAAGCGTAGATGAAGTAATTGAACGCAACGC 4733

RESULT 25

US-10-092-880-5
 ; Sequence 5, Application US/10092880
 ; Publication No. US20020164354A1
 ; GENERAL INFORMATION:

```

; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE REFERENCE: HAEMOPHILUS
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 9171
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-092-880-5

Alignment Scores:
Pred. No.: 7,75e-07 Length: 9171
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 13 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-092-880-5 (1-9171)
QY 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49
DB 3430 AATATTTCAGGTTCAATTAAGCAGAGATTACAGCTAAGAT-----GGTAGTGATTTA 3483
QY 50 AlaIleGlySerThrGlySer-----GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
DB 3484 ACTATTGTTAAACACCAATAGTGTGTGCTACTAAT----- 3519
QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
DB 3520 ---GCCAAAAGTAACCTTT-----AACCAAGGTTAAAGAT 3552
QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108
DB 3553 TCAAAAATCTCTGTGACGGT---CACAAAGTGACACTACACAGCAAGTGAA----- 3603
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
DB 3604 ACATCCGGTAGTAATAACACACTGAAGATAGCAGTGCACATAAT---GCCGGCTTAACT 3660
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlnThrLeuGluProIle 148
DB 3661 ATCGATCAAAAATGTAAACAGTAACCAATAATTACTTCTCAAAAGCAGTGACATC 3720
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
DB 3721 -----TCTGGCACAAGTGGAGAAATTACCCTAAACACAGGTACAACCAAT 3765
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
DB 3766 AACGCACCACTGGTAACGTTGGAGATACCGCTCAACACAGGTAGT-----ATC 3813
QY 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr-----ProPheIle 203
DB 3814 CTAGGTGGAATTCAGTCCAGCTCTGGCTCTGTAAACACTTACTGCAACCGAGGCGCTCT 3873
QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
DB 3874 CTTGTAGCAATATTTCG-----GCCAACACC 3900
QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
DB 3901 GTTACTGTTACTGCAAT---AGCGGTGCATTAAACCACTTTGGCAGGCTCTACAATTAA 3957

```

```

QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
DB 3958 GGAACCGAGAGTGTAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTTCTGGT 4017
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGlyLeuLys 280
DB 4018 GGCACAGTAGAGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCCAATTCAAAATTTAA 4077
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro 300
DB 4078 GCAACAAACAGCGGAGCTTAACGTAACAAAGTGCACAGGTACAAATGGTGGTACGATTTCC 4137
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
DB 4138 GGTAAATACGGTAAAT-----GTTACGGCAACCGCTGCG 4170
QY 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla 340
DB 4171 GATTTAACAGATTGGGATGCGGCAGAA-----ATTAAATCCGACAGAA 4212
QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
DB 4213 GGAGCTGCACCTTAACTACATCATCGGCAATTA-----ACT 4251
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp 380
DB 4252 ACCGAAGCTAGTTACACATTACTTCAGCCAAGGTCAGTAAATCTTTTCAGCTCAGGAT 4311
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
DB 4312 GGTACGTTGCGAGAGTATTAAATGCCCAATGTGACACTAAATVACTACAGGCACTTTA 4371
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
DB 4372 ACTACCGTGAAGGGTTCAAACATTAAATCAACACGCGGTACTTGGTTATTAAACGCAAA 4431
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426
DB 4432 GACGCTGAGTAAATGGCGCAGCATTTGGTAAACACACAGTGGTAAATGCAACCAACGCA 4491
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
DB 4492 AATGCTCCGCGCAGCGTAATCGCGACCACTCAAGC-----AGAGTGAACATCACT 4542
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
DB 4543 GGG-----GATTTAATCACAATAAATCGATTA-----AATATC---ATT 4578
QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
DB 4579 TCMAAAACCGTATTAACACCGCTACTGTTA---AAAGCGTTAAATGTGTGAATATC 4635
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLysArg 502
DB 4636 ATTCAACCGGTATAGCAAGCGTAGATCAAGTAAATTTGAAGCGAAACGC 4683

RESULT 26
US-10-282-122A-41760
; Sequence 41760, Application US/10282122A
; Publication No. US200400029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.

```


APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIORITY FILING DATE: 2003-02-20
PRIORITY FILING DATE: 2003-02-20
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-09-06
PRIORITY FILING DATE: 2000-09-06
PRIORITY FILING DATE: 2000-09-09
PRIORITY FILING DATE: 2000-09-09
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2001-02-09
PRIORITY FILING DATE: 2001-02-09
PRIORITY FILING DATE: 2001-02-16
PRIORITY FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 41760
LENGTH: 11118
TYPE: DNA
ORGANISM: Yersinia pestis
US-10-282-122A-41760

Alignment Scores:
Pred. No.: 1,3e-06 Length: 11118
Score: 172.50 Matches: 119
Percent Similarity: 32.30% Conservatives: 58
Best Local Similarity: 21.72% Mismatches: 180
Query Match: 6.67% Indels: 191
DB: 16 Gaps: 24
US-10-009-823A-1 (1-502) x US-10-282-122A-41760 (1-11118)
QY 3 GlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22
DB 7075 GGGGAATTGATCTCGGGGCGAATGGGCTTTCCGGCCAGACGTCATTACTGTAATATCGCC 7134
QY 23 SerAsnAsnLeuAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGln 42
DB 7135 AGTGGG-----GCCAGTGGCAATATTAAATGGCTATCGCCAG----- 7170
QY 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAla 62
DB 7171 -----ACGGTGGGGCGGTGACCAATACCGGTACGGTAACGTTG 7209
QY 63 GlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluPro 82
DB 7210 GGCACCGTGGGAGTTAACAGTACTGACACCTTATCATCAATACCGAATGATTATG 7269
QY 83 GlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGlu 102
DB 7270 ACCGATGGCATCTGTAATCTGGAGATGGGGGCTTCTAGCATTAGCGCGCTTA--- 7326
QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122
DB 7327 -----ACGGCAACCGTATCTCGAAT 7347
QY 123 AspProSerGly---PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLys 141
DB 7348 ATCAAGGGTGGCATTTTACCATC-----AGCATCGATAACAAT----- 7386
QY 142 LysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerPro 161

7386 ----- 7386
QY 162 AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
DB -----GGTCTGGCGGGCAACCAATATTTCCGATGGTGCATCATGCTACT 7431
QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
DB -----CTTGATTAATGGGGGACC 7449
QY 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGly 221
DB -----ATTATCGGAACCGGTAAT----- 7467
QY 222 AsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThr 241
DB -----TTGGGACGAGCGCTT 7482
QY 242 PheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSer----- 258
DB -----ATTGATGTGCTGGGGATTAACCTGTCGCGGATAATTCACTGGCTAACGTCATCAGT 7542
QY 259 -----GlyThr-----AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
DB -----GGTGACGGGACGATTAATACCAACAGACAGTACGCTGACGGGAATAGCAGCTTTAGT 7602
QY 275 -----SerAsnGlyGluLeu-----LysAsnMetThr 283
DB -----GGTGACATCATGATTGGTACCAATGGCACTGACCGTGGGTGAGCCAGTAATCTGGG 7662
QY 284 AlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeu 303
DB -----AGCAGCGCCACCGTTAATCTGGGACCCCTTACTTCTCATCTGATC 7710
QY 304 ValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThr 323
DB -----TTGATGGCGTGTAGCGAGATATTGCCAAC-----GTTCTGACGGGTGTGGCGGTTCAACA 7767
QY 324 LeuAsp----- 325
DB -----GTGATATTATCGCGGAGCAGATACCGCACTGACGGCCAATAACAGCGGCTTCTCGTGC 7827
QY 326 -----PheGlyIleLysSerGlnGlnAsnMetTrpAla 336
DB -----CAATATGCTTGGCGGTAAACAGCAACTGACGCTGGCTCAACGAACTCTG----- 7881
QY 337 GlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMet 356
DB -----GGGGCGTCTATCCAGCGTGGCGCTGCGAGGAGC-----GGCGATACCCCTCGCTGATC 7935
QY 357 ProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArg 376
DB -----GGCTTTAACGGCACCTTTGGTAAACAGCGTCAACCGGACGGGTGACTGCAAGTGAAC----- 7992
QY 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGly 396
DB -----GATGATGCCGAAGTCACTCTGACCGAGCAAC 8025
QY 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu----- 415
DB -----AGCAACGCGGTAATACTATTGATATCGCTGATCGCAGCGCTAAT 8073
QY 416 -----AlaArgPheThr-----SerGluAspGlyLeuArgArgGluGly 428
DB -----CTGGACGATATTGCTCTCTTTAATCATCGCTGACCGGTAAACGGCTTGTCTGAATGTGCG 8133
QY 429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly---LeuProGly 447
DB -----AAAAACGATGCGACGCGCTTTGACTTTGCTGCTACCGTGGCGGGCGCTTCAACCGCG 8193
QY 448 ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSer 467

QY 112 AsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGly 131
 Db 4102 AACTGAGCTTCACGCGCGCTCGCAACTCGCCCAACGCGACCGTGTCAACGTGACCGCG 4161
 QY 132 SerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAsp 151
 Db 4162 AGCAGCGCGCGCAATACCACTTCCCGCTACCAAG-----ACGGTGGAT 4209
 QY 152 PheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaVal 171
 Db 4210 TCCTCGCTGCGTCCGTCGCGAGGTGGATCCGACCAACGGTTCGGTGTATCAGCGCAC 4269
 QY 172 ValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeu 191
 Db 4270 CGCAGCGCGCAACCACTCATCATCAGCGACCGCAACCGGATTCGGCCAGGTC 4329
 QY 192 LeuGluSerTrpLysGlyAsnGly-----ThrProIleSerThrSerAsnTyrSer 209
 Db 4330 ACCGCGCAGCGACGCGCACTGCTCTTCACTCCAGGATCCGCTCGCGGATGCGACG 4389
 QY 210 TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr 229
 Db 4390 GTGTCACGTGTGGCGCGCACCCAGCAAGCATGTGCAGTCGCGCGGGGTGATCAT 4449
 QY 230 PheAspGly---AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet 248
 Db 4450 GTGATGCGTGGCGCGCGCGCGCG-----GTGATC 4482
 QY 249 AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuMet--- 267
 Db 4483 GATCCGAGC---AACGGCACCGAGATAAGCGGTACCGCGAGCGCGCGCGCGGTGATC 4539
 QY 268 -----SerGlyThrMet 271
 Db 4540 CTCACGATGGCGCGCAACCGATCGCGCGAGCGCACCGCGCGCGCAACTGG 4599
 QY 272 ThrPheSer-----SerAsnGlyGluLeuLysAsnMetThrAlaPheThr 286
 Db 4600 ACCTTACCCCGACGACCCCGCTGGCGCAACCGCGTGTATCAACCGCGTGGCGCGAGAC 4659
 QY 287 ProThrGly-----SerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
 Db 4660 CCGCGCGCAATACCGAGCGTGGCGCGAGCGTCACTGTCATCGCGCGAAGCGCGGCG 4716
 QY 301 AlaProLeuValAsn-----GlyLeuProGlnPheSerAla 312
 Db 4717 GCGCGGTGTATCAATCCGCAATGGATGCTCATCAGCGGTACGCGCGAAGCGCGGCG 4776
 QY 313 AsnPheVal-----GlyAlaGlyIleGlnProLeuThrLeuAspPheGly 327
 Db 4777 ACGGTGATCTCCCGACGCAACCGGATCGCGCGAGTCAACCGCGAC----- 4830
 QY 328 IleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThr 347
 Db 4831 -----GCGACGCGCAACTGGAGCTTCACGCGCGCGCACCGCTGGCGCAACGCGTGC 4881
 QY 348 AspIleGlyLysLeu----- 352
 Db 4882 GTGATCAATCGCTGGCGCGCGAGCGCGCGCAACACAGCAGTCCACCGCGCGCAC 4941
 QY 353 -----ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla 367
 Db 4942 GTGACTCGTGGCGCGCAGCGCGCGGTGTATCGATCCGAGCAACGGTAGCGTATCGCC 5001
 QY 368 ArgAsnGlySerSerThrArgArgTyrSerGlnAsp-----GlyTyrProGlnGly 395
 Db 5002 GGTACCGCGCGAGGTGGTGGCGCGCGTATCTCACCAGCGCGCAACCGCGATCGCG 5061
 QY 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
 Db 5062 CAG-----GTCACCGCGCGATGGC-----AGCGGCAACTGGAGCTTCACG 5100

QY 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 425
 Db 5101 CCGGCGACGCGCTGTCCATGTCACGCGTGTCAATGCGTGGCCGACGAGCT----- 5154
 QY 426 ArgGluGlyAsnAsnHis-----TyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
 Db 5155 ---GCGGCAACACCGCGCGCGCTCAGCACCGACCGTGTGACGC---GGTGGCCCC- 5204
 QY 444 GlyLeuProGly 447
 Db 5205 GCGCACCGCGGT 5216

RESULT 29
 US-10-282-122A-30151
 ; Sequence 30151, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zvakind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA 034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 30151
 ; LENGTH: 7407
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-10-282-122A-30151

Alignment Scores:
 Pred. No.: 3,466-06 Length: 7407
 Score: 166.00 Matches: 113
 Percent Similarity: 25.12% Conservative: 64
 Best Local Similarity: 22.42% Mismatches: 189
 Query Match: 6.42% Indels: 139
 DB: 16 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-282-122A-30151 (1-7407)

QY 20 GlyThrValSerAsnAsnIleAlaAsnThrIleGlyTyrLysGlnGlnVal 39

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...372
SEQUENCE DESCRIPTION: SEQ ID NO: 77:
US-10-335-977-77

Alignment Scores:
Pred. No.: 5.39e-08 Length: 372
Score: 165.50 Matches: 39
Percent Similarity: 51.64% Conservative: 24
Best Local Similarity: 31.97% Mismatches: 48
Query Match: 6.40% Indels: 11
DB: 16 Gaps: 3

US-10-009-823A-1 (1-502) x US-10-335-977-77 (1-372)

QY 392 IleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyr 411
Db 7 ATTGAGAAATCGGTGATTTCTTAGCCTTTAGTAACGGGTGTAGAGCGCGTTCGCT 66
QY 412 AsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnHis 431
Db 67 CGCATCGGTATTWAGCTTTCACTAAGCATCAAGGCTTAAGGAAATCGCGGTAACTTC 126
QY 432 Tyr-----SerAlaThrLeuAsp-----SerGlyGlyProGlu 442
Db 127 TATGAATGCAAGAGCCATTAATGCGGAAACAGACCCCTAAGSGGTAAACCCAT 186
QY 443 PheGlyLeu-----ProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
Db 187 TTAGGTGGGACGAGAGGCGCAAGCTCAAGTTTGGGAAATCAGGCACAAATATTAGAA 246
QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe 480
Db 247 ACAGCAACCTGATGATCGCGGAAACGCCCTAACCAATCTCATTTAATGCAAGAGCTAT 306
QY 481 GlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGlu 500
Db 307 TCTATGACGCTAGAGCTTTGGCGGCGGATGACATGATCAAGAGCAATAGCTTG 366
QY 501 LysArg 502
Db 367 AAAAAA 372

RESULT 31

US-10-114-170-39

Sequence 39, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESS: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,170

FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 25165
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-114-170-39

Alignment Scores:

Pred. No.: 2.53e-05 Length: 25165
Score: 165.00 Matches: 145
Percent Similarity: 33.70% Conservative: 71
Best Local Similarity: 22.62% Mismatches: 189
Query Match: 6.38% Indels: 236
DB: 14 Gaps: 35

US-10-009-823A-1 (1-502) x US-10-114-170-39 (1-25165)

QY 8 GlyAlaThrGlyMetLysThrHis-----SerThrGlyLeuGlyThrValSer 23
Db 13018 GGAACAGCGGAGCGGTACCCATCAGGTGAGGTGCAATACCGGCTCCGACCATACC 13077
QY 24 -----AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGln 37
Db 13078 TTAAAGCGCATCAGCGGCGATAACATCTGAAACCGCGATGAAAGGC----- 13125
QY 38 GlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIle-----GlySerThrGly 55
Db 13126 -----CAGCGGTGACCATCAGCGGCGGCGAGTACGCGG 13158
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 13159 -----CTGGCGAGCGGCGCGAGGTC-----ACGTCAGC 13188
QY 76 ThrGlnGly-----AlaPheGluProGlyAsn----- 84
Db 13189 CTCACGGTCACTACAGCGCCACCGACGCGCATCGGCACTGGACCTTAACCGTG 13248
QY 85 SerValThrAspLeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLys 104
Db 13249 CCGGTGAGCGATCTGCGGCATTAGGTGAGGCACTATACGTCAGCGCCAGCCGCCACC 13308
QY 105 ValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspPro 124
Db 13309 -----AGTGCAGCAGGCAACCGCCAGCAGCGCGCAATTACTGTCGAC--- 13356
QY 125 SerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThr 144
Db 13357 -----ACGCGCTGCGCGAGCTCACCATCAACACC 13386
QY 145 LeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThr 164
Db 13387 GTGGCA-----GGCAGCATATTATCAACCGCCGCGGAGCGGGGCC 13428
QY 165 SerThrAlaLeuAsnAlaValValAsn-----LeuGlyAspSerThrAspLysThr 181
Db 13429 GATCAAAACCATCAGCGGCGGTGTGACTCGCGCGCGCTGGCGATACGTCACCGTACG 13489

QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
Db 13489 CTGGGC---GGGAACACTTACACCGCTACGGTACAG----- 13521
QY 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnP--- 220
Db 13522 -----AGCACTTAACTTGGAGCTGACGGTTCGACAGCGGATCTCCAGCG 13569
QY 221 ---GlyAsnSerHis---AspIleThrValTyrPheAspGlyAlaProSerSerThrGly 238
Db 13570 TTGGCAATGTGTGATTGACCATTTACCGCTCGGTCAACCAACGCTAATGGCAACACCGGG 13629
QY 239 SerLysThrPheGluTyrLeuVal----- 246
Db 13630 AGCGCACGCGGATATCACCATGTATGCCAACCTGCCGGGGCTGGCGGTAGATACCGTG 13689
QY 247 -----AlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThr---Asp 261
Db 13690 GCGGCGATGATATCGTCAACAGCATCGAGCAGCGAGCGCTGGTGTATCACCAGCGGC 13749
QY 262 SerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn 281
Db 13750 AGTACGCGCCTGAATGACAGGTGTGTGTGCTGACGGTTACCATCAACAGT----- 13797
QY 282 MetThrAlaPheThrProThr-----GlySerAlaThr 292
Db 13798 ---GTGGGTATTCGCGCACCGTCGAGCGGACGAGCTGGAGCTGGCATTCGCGCG 13854
QY 293 LysAspLeuAlaTrpGlnProAlaProLeu-----ValAsnGlyLeuProGln 309
Db 13855 GCAACAGTCACGCGCTCGCGGCGGCTTAAACCGTGGAGGTAGACGCG---CAA 13908
QY 310 PheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe----- 326
Db 13909 AGCAGCGCCATTAACCCAGTCAGCGTCAGCATCCGTTCCCGTCAATTAACGGCGGTG 13968
QY 326 ----- 326
Db 13969 GCAATCAGCATCAACACCGTTCCAGCGACGACGTGATTAAACGCGCAGAAAAAGGCACC 14028
QY 327 -----GlyIleLysSerGlnGlnAsnMet----- 334
Db 14029 AATCTGACTCTTTCGCGCAGTACAGCGGATTTAGAGCGGCAAAACCGTCACCGTCACT 14088
QY 335 TrpAlaGlyAla-----ProAlaSerAlaAlaIleGlyThr----- 347
Db 14089 TTTGGCGTAAACCTACACTGCAAGCGTCGCGCAACCGGAGCTGGAGTGTAAACGTT 14148
QY 348 -----AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
Db 14149 CCGCGCGAGATCTGGCAACTCTGCCAGAGGCGCGCGGAAATGTGACGCGCAGCGTTAGC 14208
QY 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp----- 380
Db 14209 AGCGGAGCGGTACAGTGCCTCGCGACCCATCGGTATAGCTTGACGCCAGCGCGCG 14268
QY 381 -----GlyTyr 382
Db 14269 ACGTCAACCATTAACACCATCGCGCAGCATATCTTAAACCGCAGAACCGGAAGC 14328
QY 383 Pro-----GlnGlyAspLeuValAspValThrIle 392
Db 14329 CCGCTCACCATCAGCGCACACGACCGCGAAACCGGCGAGAGCGGTGACCGTCAACCTT 14388
QY 393 -----ThrSerGluGlyLysLeuGln-----GlyLysTyrSerAsnSerGlnVal 407
Db 14389 AACGGCGCAACCTACACCGCACTGTGACGCGGACGCTAGCTGGAGCGTCAAC----- 14442
QY 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427
Db 14443 -----GTTCC-----ACTTCAGCCCTGGCGCGGTCAACGCA 14475

QY 428 GlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly 447
Db 14476 AGCAATTACACCGTCCAGCCACGGTCAATGAC-----AAAGCGGCAACCCCGGC 14526
QY 448 ThrSerAsnTyrGly-----LysLeuSerValAsnGlnLeu 459
Db 14527 AGCGCAGCATTAATCTGGCGGTAGACACCGCGCGCGGTCTCCATTAAACCGGTG 14586
QY 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArg--- 478
Db 14587 GCGGCGGATGACATCATCAACGATCGCAACATGCGGAGCGCTGCTGATCTCCGCGCAC 14646
QY 479 -----GlyPheGlnMetAsnSerLysValThr 488
Db 14647 AGTAGCGCGGGAAGCGGCGCATGTGTGAGCGTGTGCTCAACGCGCAACCTACACC 14706
QY 489 Thr 489
Db 14707 ACC 14709
RESULT 32
US-10-335-977-57
; Sequence 57, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1029
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-10-335-977-57

Alignment Scores:

Pred. No.: 3.31e-07 Length: 1029
Score: 164.00 Matches: 84
Percent Similarity: 39.58% Conservative: 49
Best Local Similarity: 25.00% Mismatches: 130
Query Match: 6.34% Indels: 74
DB: 14 Gaps: 14

US-10-009-823A-1 (1-502) x US-10-335-977-57 (1-1029)

QY 32 IleGlyThrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeuAlaIle 51
Db 7 CTGGGCTTAAAGTATTCTAGGCTCTTTTGTGGATATGCTCTCAAGTCAAACTCATC 66

QY 52 GlySer-----ThrGlySerGlnGlyProAsn-----GlnAlaGlyMetGly 65
Db 67 GTACACGCCCTTATAAAACGGGTGGCGGCGAGAAATGACTTTCTGTGGGGCTTGG 126

QY 66 AlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSer 85
Db 127 GTAGCGTGTGATCGACGACTAAATCTTTTCGAAGCAATATCCAAACACGGATGTC 186

QY 86 ValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysVal 105
Db 187 AAAACCGCATCTAGCATTCAGGCGATGGCTTTTATCATTAGCCCTGATAGGGGATC 246

QY 106 -----HisTyrThrArgAla-GlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAs 123
Db 247 ACGGCAATTCACGAGCATGGGAGTTCTTTGGACTCCAGGCGATTTGGTTAC 306

QY 123 pProSerGlyPheThrLeuMetGlySer-----ArgIle 134
Db 307 CACGGTGGCTTGTGTCAGATCGGTGAGAAACAGGAGCATACCGGTAAATAAGGG 366

QY 134 eSerAsnAsnProAsnIleLysGlyThrLeuGluProIleGlnLeuAspPheAsnAs 154
Db 367 -AGGATACGGAGCTTTAAAGTGGATATACCGGCCCTTTAGAAAACATTAGGATGA 425

QY 154 pProThrValAlaLysSerProAlaLysThrSerThr-----AlaLeuAsnAlaValVa 172
Db 426 TCAGGAATGTG---ATGCCTGCTAAACAGTAAACCGCATTTCTAGGCGGCAATAT 482

QY 172 LAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeu 192
Db 483 AAACGCTGGAAGCATCGCGATCAGACGCGCG-----GTATTCGCTTTAGA 530

QY 192 uGluSerTyrLys-----GlyAsnGlyThrProProIleSerThrSerAsnTyrSerTy 210
Db 531 TTCTTACGCAAAACCCCTTCAGATGCGATTATCCAGTGTATGATTCAGGCAGATCT 590

QY 210 rAlaGlnProMetArg-----ValTyrAspGlnGlnGlyAsn----- 222
Db 591 AGCTCACGTCGCCGAAACATGGGATCTTTATACATGAAGATGGCGAGCTCTTTATT 650

QY 223 -----SerHisAspIleThrValTyrPheAspGlyAla----- 233
Db 651 GATTGAAACCATCGGATTTGGGTGAGCTATAGAGTGGGAAATGTTAAGACATCCT 710

QY 234 -ProSer----- 235
Db 711 CCCTTCTGCAGAAACAGACGCTTGAGTTGAATGGGGTTAAGATTTCTTTCAGAACGA 770

QY 236 -SerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySe 255
Db 771 TTCAGCGGTGACGAGGACTCAAGCTTAGTGGCGGTAAATAATGCCATCATCGGTTAA 830

QY 255 rAlaAlaSerGly-----ThrAspSerAlaGlyLeuLeuMetSerGlyThrMe 271
Db 831 AAGCCAAACAGGCATTGAAGCTTATTTGGAGCGGCAAGCAATTCGTTTGGAAACACCAA 890

QY 271 tThrPheSerSerAsnGlyGluLeuLysAsnMetThr-----Al 284
Db 891 CGAATTAGCGCGCATGAAAGCTTAAATAACATTTGTGGTCACTCAAGCCGCGGAGC 950

QY 284 aPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGln 299
Db 951 GTTCGCTAACTTTTGTAGACGCGGATAAAGATGTAAACGGCTTTCAAA 996

RESULT 33
US-10-147-299A-3
; Sequence 3, Application US/10147299A
; Publication No. US20040058323A1
; GENERAL INFORMATION:
; APPLICANT: KO, ALBERT I.
; APPLICANT: HAAKE, DAVID A.
; APPLICANT: REIS, MITERMAYER GALVAO
; APPLICANT: MATSUNAGA, JAMES
; APPLICANT: CRODA, JULIO HENRIQUE ROSA
; APPLICANT: SIQUEIRA, ISADORA CRISTINA
; APPLICANT: KILEY, LEE W.
; APPLICANT: BAROCCHI, MICHELE
; APPLICANT: YOUNG, TRACY ANN
; TITLE OF INVENTION: PROTEINS WITH REPETITIVE BACTERIAL-IG-LIKE (BIG)
; FILE REFERENCE: 3673-19
; CURRENT APPLICATION NUMBER: US/10/147,299A
; CURRENT FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 5863
; TYPE: DNA
; ORGANISM: Leptospira kirschneri
US-10-147-299A-3

Alignment Scores:
Pred. No.: 6.47e-06 Length: 5863
Score: 162.00 Matches: 125
Percent Similarity: 31.73% Conservative: 80
Best Local Similarity: 19.35% Mismatches: 241
Query Match: 6.26% Indels: 200
DB: 16 Gaps: 23

US-10-009-823A-1 (1-502) x US-10-147-299A-3 (1-5863)

QY 7 IleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly---ThrValSerAsnAsn 25
Db 1229 GTGTGTAGTAGTAATATCATCCGCGCCCATCGGTGGAGTAGATATACGGTTTCTTAAT 1288

QY 26 IleAlaAsnAla----- 29
Db 1289 GTTACCAACGCCACTTTTGAATTCGATTCAAGTGGTTCCGATTCCCATTCGATAGTCGA 1348

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 1349 GGTACGTCTACGTTTGTACAAGCGGATAGGAGTCTACTCGGACGGTTCTTCTCAAAACA 1408

QY 49 ----- 49
Db 1409 AGTGATCAAGTGCCTGGACAGCTTAATTTCTTCAATATTACAAATATCTAATTAAT 1468

QY 50 AlarIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 1469 GCAGTTCCTCCAAAGAGAAATAAATCTCTCTTCTCCGGAGCGCTAGGT----- 1516

QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 1517 -----ACAGCAAGGATCACCGCACTTTAGAGCAATCTCCTCATATACCGACATC 1567

QY 90 AlarIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArg 109
Db 1568 TCGGTCAATGACGAACTTTTATGTTTCTATCGAAGTGTCCACCCACAAATCTTCGTATCT 1627

QY 110 AlaGly---AsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 1628 TCAGGACTTACCGTTCTTTCACGGCGGACCGAGTTTATACGGATCGAATCAATCAAAAT 1687

QY 129 LeuMetGlySerArgIleSerAsnProAsnIleLysLysGluThrLeu----- 145
DB 1688 CTGACTTCTCAAGTAACCTTGAATCTCCCAACGAGACGAGTACAAATCAGCAACGCA 1747
QY 146 -----GluProIleGlnLeuAspPheAsnAspProThr----- 156
DB 1748 AACGGAACCTCAAGGAATGCTTGGCTCTCTGTCGGAACCTACGAACATATCAGCAACG 1807
QY 157 -----ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn--- 173
DB 1808 TTAGTGGGTACTTCTTCGCTACCACTCTTACGCTCAAAACGCGGTCTTAAATTCG 1867
QY 174 -----LeuGlyAspSerThrAspLysThr--- 181
DB 1868 ATCAGATTACTCCGCTCTCTTCCTTCGTCAGTACGAGGAGAACTCTGAACCTTACTGCA 1927
QY 182 -----GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGly 197
DB 1928 ACCGGAACCTTATCTGACGGAAGTAACCAAGATTAACTACCTCCGTCGCTTGGACGAGT 1987
QY 198 AsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyr 217
DB 1988 ACGGATCTTCATCGTCTTCGTAGACACGCTCAGTAGACAGCGGCGACAGCAGGT 2047
QY 218 AspGlnGlnGlyAsnSerHis----- 224
DB 2048 GTTGCACAAGTAACACTCAGATCAGTCAGTCACATTTAGCGGAACTTCTTCTGTATCAAT 2107
QY 225 -----AspIleThrValTyrPheAspGlyAlaProSer 235
DB 2108 TTTACGTAAGTACGAGCGGTTTATAGATTCAATTCAGTAATCTCGAAGATTCTCCGATT 2167
QY 236 SerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySer 255
DB 2168 GCAAAAGGAACCTCTACA-----AGAGCAATCGCGAGGTGTTTTTTCAGACGGAAGC 2221
QY 256 AlaAlaSerGlyThrAspSerAla----- 263
DB 2222 AATTGATATATAGTATGATCAAGTTATTGGATAGTTTCAAAACAAACGATGATCCAGCTA 2281
QY 264 GlyLeuLeuMetSerGly-----ThrMetThrPheSerSerAsnGlyGluLeu 279
DB 2282 GGAGTTTGAACACCGGCTCTAAAGAACTGATGAATCTCCGCAATGCA----- 2335
QY 280 LysAsnMetThrAlaPheThrProThrGlySerAlaThr----- 292
DB 2336 ---AACAGTACCTGGAACCTCAAGGATCACTGCAACGTTAGGAGGTGTGACGGGATAC 2392
QY 293 LysAspLeuAsnAlaTyrGlnProAla----- 301
DB 2393 GCCGATCTTACAGTAATCGCTCCAGTTTAAACAGCATTCAAATCGATCTACACATCCG 2452
QY 302 ProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnPro 321
DB 2453 AGCGTTGCAACGCTCTGACTCAA-----AATTTTACTGCAACCGGAGTT----- 2497
QY 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet-----TrpAla 336
DB 2498 -----TACTCAGATGGTAGCAATCAAGATCTAACCGATTCCGTTACTTGGCG 2545
QY 337 GlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMet 356
DB 2546 TCTTCCAATCTGCTGTTCGACGATCAGCAACGCTTCGCGAACAACGGAAGTACT 2605
QY 357 ProIleGlnThrSerSerThrAlaAsnGlySerSerSerThrArgArg 376
DB 2606 ACTCTTCAAACTGATCCCAATATCAGCGGAGTCTGGCGCCCACTACTTCT----- 2659
QY 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGly 396
DB 2660 -----GATCCAAGTGTATTAAACGGTTACA----- 2683
QY 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla 416

DB 2684 -----AACGCAACCTTAAACAAGTATCAGCATCGCTCCACCTCT 2722
QY 417 ArgPheThrSerGluAspGlyLeuArgGlu----- 427
DB 2723 TCTTTCACATCGCAAAAGGATTAATCAAGACTTTGTAGCGACCGGTTATTATACAGAT 2782
QY 428 -----GlyAsnAsnHisTyrSerAla 434
DB 2783 GGTTCCTTCTAGAGACTGACCACTCAAGTCACTTGGAAATCTTCCATACTTCTACCGCT 2842
QY 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
DB 2843 ACGATCAGCAATGCAACGGAACCTCAAGGAAGAATGGCGGTCGATCTGTTCTACA 2902
QY 455 SerValAsn-----GlnLeuGluThrSerAsnValAspMetSerArg 468
DB 2903 AATATCTCCGCTCTTTAGAGGAACGATATAGTCAGACCAAAACGTAACCGTTACATCT 2962
QY 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
DB 2963 GCGGTTCTGAATTCGATCCAGTTTCTCCAGCGGACATTAGTGTAGCCAAAGGAACACC 3022
QY 489 ThrAlaAspThrMetLeu 494
DB 3023 AAGGCTTACCGCGATC 3040
RESULT 34
US-10-282-122A-22709
Sequence 22709, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282.122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22709
LENGTH: 3693

TYPE: DNA
ORGANISM: Helicobacter pylori
US-10-282-122A-22709

Alignment Scores:
Pred. No.: 5,4e-06 Length: 3693
Score: 160.00 Matches: 115
Percent Similarity: 35.93% Conservative: 74
Best Local Similarity: 21.86% Mismatches: 179
Query Match: 6.19% Indels: 158
DB: 16 Gaps: 26

US-10-009-823A-1 (1-502) x US-10-282-122A-22709 (1-3693)

```
QY 20 GlyThrValSerAsnAsnIleAlaAsnThrIleGlyTyrLysGlnGlnVal 39
Db 1624 GGGACTACCACTAATACTAAGCTAAAGCAAGCTCCCAACTAAAGCGATGGTGG 1683
QY 40 Val-----PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer---ThrGlySer 56
Db 1684 GTGAATAATGAAGAAGAGCTAAAGCGCAATTTAGCCAAAGCAGCGGGAACCAACA 1743
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 1744 CAATCTCCTAACAGCAGCGGTGATGGAGCT-----TTAAACACCGGTGTG--- 1788
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
Db 1789 -----CAAAATGTCAGCAAT----- 1803
QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThr 116
Db 1804 TTCACAACAAAGCATT----- 1818
QY 117 GlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsn 136
Db 1819 CAACACCTTTTCAAAACCAAGAAAGTAATATCCAGCTTGGCGGAATGCAATTTATAAC 1878
QY 137 AsnProAsnIleLysGlyLysGluThrLeuGluPro-----IleGln 149
Db 1879 ACTAATGGGAGTCAGTCGCAAGAGATGACACCTAACCAATAACCAAGATTTCAGCATCCAA 1938
QY 150 LeuAspPheAsnAspProThrValAlaLysSer---ProAlaLysThrSerThrAlaLeu 168
Db 1939 TTGAGGCGGAATTTTACCAGCTCATCAATCAATTTAACAGCAAGTGCCTACAGCATG 1998
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 1999 AATGCTTTAATATCAAGCCCAACCAACCAACCAAGCGGATCAGCAAGCAATAAT 2058
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyr 208
Db 2059 AACGCATGCGCGAGTGAATGAGTGGG-----AGTAATGGTAACCTGG 2100
QY 209 SerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrVal 228
Db 2101 TGCTATACGCAA-----TGTCGGATTCTAGGCT 2130
QY 229 TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuAlaMet 248
Db 2131 TATTACAGCGGTTGCAAGCGCTTTAGGTATCAAAAG----- 2169
QY 249 AsnProSerGluAspGlySerAlaAspSerGlyThrAspSerAlaGlyLeuLeuMetSer 268
Db 2170 -----CAAGCGCACTCAAGCGGAGCAATGTGGGAACAGCATCACCTAC 2217
QY 269 GlyThr-----MetThrPheSerSerAsnGlyGluLeuLys----- 280
Db 2218 AATGTCACCAAAATCAGCTCACTAGTAATGGTTGCTCAACCAAAATCATCAAAATCTT 2277
QY 281 -----AsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294
Db 2278 AAGACGGTTAATGGAGGCAATGGCGCGAGTGGTACAGGCAGTGGGAATGGCCAGCTCAA 2337
```

```
QY 295 LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
Db 2338 ATCAAC----- 2343
QY 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
Db 2344 -----ACAGCCTACAGATGCTCACA----- 2364
QY 335 TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
Db 2365 -----GAGCGCAGCGATGGGAATTAGGACT----- 2391
QY 355 MetMetProIleGlnThrSerSerGly-----AsnSerThr 366
Db 2392 -----TATAGTAGTAGTAGTGGCAGTAATAACGGCTATACGCCCATGCAATAGCACC 2442
QY 367 AlaArgGlnGlySerSerThrArg---ArgTyrSerGlnAspGlyTyrProGlnGly 385
Db 2443 -----AATGGAGCAATAAACAAGAGTGGGACAAATTGTTATGAACCAACAAACAA 2496
QY 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
Db 2497 AAGCGCACCCAGCGCAACCGCCACACAGCAATTTACAAAAGTCTATATATGACGCC 2556
QY 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer-----GluAsp 422
Db 2557 CAAAAATA-----GCCAACATTATCCCGAGCTCTGGGAACAATAAAGCGGTGAAAC 2610
QY 423 GlyLeuArgArg-----GluGlyAsnAsnHisTyrSerAlaThrLeuAspSer----- 438
Db 2611 GCTTAAACAATCTTTGAAGCGTTAAAAATAATAGCAGCAGTCTCAGTAATTTATGT 2670
QY 439 GlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu-----SerVal 456
Db 2671 GGTAAATGTTAGTAGCGGTAGTAGTGGCACTACTTCTCGGTGGCTTATCAACCTTTA 2730
QY 457 AsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle----- 474
Db 2731 GGGGCAATCCCAACCAATGAGTGGAGCGATAGCAATAATTTAATTAATCTGCTCACTGAA 2790
QY 475 -----IleIleGlnArgGlyPhe---GlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 2791 TTCAATTAACCGCGGGTTTATCCAAAATAATAGTAGTAGTGTATCTACTAGTCTTACA 2850
QY 493 MetLeuGlnLysAlaLeu 498
Db 2851 AGCGCTTTTCAAGCCATT 2868
RESULT 35
US-09-881-752A-149
Sequence 149, Application US/09881752A
Patent No. US20020115078A1
GENERAL INFORMATION:
APPLICANT: Kleanthous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Genome
FILE REFERENCE: 06132/041002
CURRENT APPLICATION NUMBER: US/09/881,752A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/833,457
PRIOR FILING DATE: 1997-04-01
NUMBER OF SEQ ID NOS: 370
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 149
TYPE: DNA
ORGANISM: Helicobacter pylori
```

FEATURE:
 NAME/KEY: CDS
 LOCATION: (51)...(3740)
 US-09-881-752A-149

Alignment Scores:

Pred. No.: 5,618-06 Length: 3793
 Score: 160.00 Matches: 115
 Percent Similarity: 35.93% Conservative: 74
 Best Local Similarity: 21.86% Mismatches: 179
 Query Match: 6.19% Indels: 158
 DB: 26 Gaps: 26

US-10-009-823A-1 (1-502) x US-09-881-752A-149 (1-3793)

QY 20 GlyThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnVal 39
 DB 1674 GGGACTACCACTAATPACTCAAGCTAAAGCAACGCTCCCAACTAAAGCGATGGTGTG 1733
 QY 40 Val-----PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer---ThrGlySer 56
 DB 1734 GTGAATATGAGAGAGAGATAAACGGCCCAATTAGCCCAAGCAGCGGAGCAACACCA 1793
 QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
 DB 1794 CAATCTCTAAACAGCAGCGTGTGGAGCT-----TTAAACACCGGTGTG--- 1838
 QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysPhe 96
 DB 1839 -----CAAAATGTCAGCAAT----- 1853
 QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThr 116
 DB 1854 TTCCAACAAGCAAT----- 1868
 QY 117 GlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsn 136
 DB 1869 CAAACGCTTTTCAAAACCAAGAAAGTAATATCAAGCTTGGGCGAATGCGATTATAAC 1928
 QY 137 AsnProAsnIleLysLysGluThrLeuGluPro-----IleGln 149
 DB 1929 ACTAATGGAGTCAAGTCAAGAGATGACACCTAAACAATACCAAGATTACCACTCCAA 1988
 QY 150 LeuAspPheAsnAspProThrValAlaLysSer---ProAlaLysThrSerThrAlaLeu 168
 DB 1989 TTGAGGCGCAATTTTACCAGCTCATCAATACCAATACCAAGCAAGTGCCTACAGACATG 2048
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 DB 2049 AATGCTTTAATATCAAAACCAACCAACCAACCAACCAACCAACCAACCAACCAATAT 2108
 QY 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProPheIleSerThrSerAsnTyr 208
 DB 2109 AACCGTCGCGAGTGAATAGTGGG-----AGTAATGTTAACTCG 2150
 QY 209 SerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisaspIleThrVal 228
 DB 2151 TGCTATCAGCAA-----TGTCCTGATCTTAAGGCT 2180
 QY 229 TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet 248
 DB 2181 TATTACAGCGGTGCAACGCGCTTTAGGTATCAACG----- 2219
 QY 249 AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSer 268
 DB 2220 -----CAAGCGCAACTCAAGCGGAGCAATGGTGGGAACGACATCACTAC 2267
 QY 269 GlyThr-----MetThrPheSerSerAsnGlyGluLeuLys----- 280
 DB 2268 AATGTCACCAATCAGCTCACTAGTAATGTTGCTCAACCAATATCAACAATCTT 2327
 QY 281 -----AsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294

DB 2328 AAGAGCGTTAATGAGGCAATGCGCGAGTGTGTACAGGCAAGTGGGATGGCCACAGTCAA 2387
 QY 295 LeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
 DB 2388 ATCNAC----- 2393
 QY 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
 DB 2394 -----ACAGCCTACAGATGCTCACA----- 2414
 QY 335 TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
 DB 2415 -----GACGCCAGCATGGGAANTTAGGAGCT----- 2441
 QY 355 MetMetProIleGlnThrSerSerGly-----AsnSerThr 366
 DB 2442 -----TATAGTAGTAGTAGTGGCAATTAACGCTATACGCCATGCAATAGCACC 2492
 QY 367 AlaArgAsnGlySerSerSerThrArg---ArgTyrSerGlnAspGlyTyrProGlnGly 385
 DB 2493 -----AATGGAGCAATAAACAAGTGGGAACAATTTGTTATGAACCCCAACAAACAA 2546
 QY 386 AspLeuValAspValThrIleThrSerGlyLysLeuGlnGlyLysTyrSerAsnSer 405
 DB 2547 AACGCCACCGCAACCGCCACCAACCGACAGCAATTTACAAAAGTCTATAATGACGCC 2606
 QY 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer-----GluAsp 422
 DB 2607 CAAAAATA-----GCCAACATTATGCCAGCTCTGGGAACAATAAAGGCGTTGAATAAC 2660
 QY 423 GlyLeuArgArg-----GluGlyAsnAsnHisTyrSerAlaThrLeuAspSer----- 438
 DB 2661 GCCTTAAACAATCTTTGAAGCGTTAAATAATATAGCAGCAGCTCTCAGTAATTTATGT 2720
 QY 439 GlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu-----SerVal 456
 DB 2721 GGTATGTTAGTAGCGGTAGTAGTGCGCACTACTTGTCCCGTTGGCTTATCAACCTTTTA 2780
 QY 457 AsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle----- 474
 DB 2781 GGGCAATCCCAACCAATGGAGTAGGAGTAGGAGTAGGAGTAGGAGTAGGAGTAGGAGTAG 2840
 QY 475 ---IleIleGlnArgGlyPhe---GlnMetAsnSerLysSerValThrThrAlaAspThr 492
 DB 2841 TTCATTAACCGCGCGGTTCATCCAAAATAATGATAGTAGTAGTAGTAGTAGTAGTAGTAGT 2900
 QY 493 MetLeuGlnLysAlaLeu 498
 DB 2901 AGCGCTTTTCAAGCCATT 2918

RESULT 36

US-10-282-122A-12963
 ; Sequence 12963, Application US/10282122A
 ; Publication No, US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA 034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 PRIOR FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21

Db 8007 CTCTCGTGGATGCAAGACGTACAAAGTGTCTGCTAGACATGAGCGGTGAACGC 8066
 Qy 413 eProleuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSe 433
 Db 8067 CAAGATCGCAACG---GGATCGACGACGGCGGTGTGTACGACACGCTGGCACAATAACAA 8123
 Qy 433 rAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro----- 446
 Db 8124 GTTGACGTC-----GGCGAGTCAATGCCACACGCGGTGAGGTGCAACAGTCGC 8177
 Qy 447 ---GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSer-----As 463
 Db 8178 CGCCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 8237
 Qy 463 nValAspMetSerArgGluMetValAsnMetIleIle 475
 Db 8238 COTGATACGTCGGGTAAACGTGACCAACTCGTTCGTC 8274

RESULT 37

US-10-369-493-37512
 ; Sequence 37512, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO. 37512
 ; LENGTH: 2037
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas fluorescens
 US-10-369-493-37512

Alignment Scores:

Pred. No.:	1.1e-05	Length:	2037
Score:	153.50	Matches:	126
Percent Similarity:	31.42%	Conservative:	71
Best Local Similarity:	20.10%	Mismatches:	205
Query Match:	5.94%	Indels:	225
DB:	15	Gaps:	30

US-10-009-823a-1 (1-502) x US-10-369-493-37512 (1-2037)

Qy 5 LeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsn 24
 Db 4 CTCAATATCGGATGTCGGGTGGCGCGAGTTCATCTCTCTGCGTGTGACGGCAAC 63
 Qy 25 AsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAspLeu 44
 Db 64 AACATTCCAACTGTCGACCGCGGTATTACGCGACGCAACCGTG----- 111
 Qy 45 PheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAla----- 62
 Db 112 -----CAGGGCACCAAGTCTCGATTGAG 135
 Qy 63 -----GlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
 Db 136 TACGGTAACGTGTTATCGGTACCGGACACTGTGCGACGTCGCGCGGTGTACAAC 195
 Qy 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeu-----AlaIle 91
 Db 196 TCCTACTCTGAATCGCAGCTGCGTACCGCCACGTCGCTCAATAGCAGCGCGCTTC 255

Qy 92 GlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGly 111
 Db 256 GGTGCCAGCGG-----ACCGCGCTGGACGCTGCTGCTCGACACCAACACTGCGC 306
 Qy 112 AsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGly 131
 Db 307 CTGACCGGTGTCTGCAGAAATTCCTACC-----TCGATGCAAGGC 348
 Qy 132 SerArgIleSerAsnAsnProAsnIleLysLysGluThrLeu-----GluPro 147
 Db 349 GTGTCGACCTCGGCCACCGACGACACTTCCGCCAGTCGGTGTGACCGCGGCCAGGCC 408
 Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
 Db 409 CTGACCACTGCTTCAAT-----CGCTGGCCAGCAGATGAACGATCAGACGCCACG 462
 Qy 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
 Db 463 CTCAATGGC-----AACCTGTCCGACATGGCTCCAGGTGAACAACTGGCAACCTCG 516
 Qy 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsn 207
 Db 517 ---ATTGCCAGCTGAACACGAGATCGCGAG-----ATCTCCACCGCGGC 561
 Qy 208 TyrSerTyrAlaGlnProMetArgValTyrAsp----- 218
 Db 562 -----GGCCAGCCGACGATCTGTCGACACGCCGTAAACGAAGCGTGGTCAATC 612
 Qy 219 -----GlnGlnGlyAsnSerHisAspIle----- 226
 Db 613 TCCGAGCTGACCGCGCGCAAGTCTTGAGCGTGCACCACTTCGACATCTACATCGGC 672
 Qy 227 -----ThrValTyrPheAsp 231
 Db 673 AGCGGCCAGCGTGGTTCATCGGCAATACCAATACCCCTGAGCAGCCGCTCCGCTCAAG 732
 Qy 232 GlyAlaProSer-----SerThrGlySerLysThrPheGluTyrLeu 245
 Db 733 GGTGACCATCGCGATCGGATCGATGATCGATGATCGTGGTTCGAGCACCATCGACATCACC 792
 Qy 246 ValAlaMetAsnProSerGluAspGly----- 254
 Db 793 TCGGCGATGACCGGTGGTGAATCGCGGTCTGCTGACTTATCGCAAGAGTGTCTGAT 852
 Qy 255 -----Ser 255
 Db 853 CTGCACTCAACGAGCTGGCGCGTGTGGCACTGCTGTCGCGACCATCAACCGCCAG 912
 Qy 256 AlaAlaSerGlyThrAspSerAlaGly----- 264
 Db 913 CAAGCCAGGGATCGACACAAAGCGTGACTTTGGCGGGCCATTTCAACACATCAAC 972
 Qy 265 -----LeuLeuMetSerGlyThr----- 270
 Db 973 AGTGGCGCTGATCATGTCAGTCAGCGCAGATCGCCAGTCCGCAACAGTCGAGGTTCGGGC 1032
 Qy 271 -----MetThrPheSerSerAsnGlyGluLeu----- 279
 Db 1033 AATCTTGATGTCACATCAAGACACCGGACAGCTGACACCACCGACTATCAGGTCAAC 1092
 Qy 280 ---LysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAlaTyr 298
 Db 1093 TTCACGACGCGGACCACTACACGCTCAAGCGCTCCGACGCGCACCGACATGGGTTCGTT 1152
 Qy 299 Gln-----ProAlaProLeuValAsnGlnLeuProGlnPheSerAlaAsnPhe 314
 Db 1153 AGCACCGACCGCGCTCCGCTTAATCGACGGCTTT----- 1191
 Qy 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnAsnMet 334
 Db 1192 -----ACCTGGCCCTTAACCGTGTGCTGAGCGCGCGGATACG 1233
 Qy 335 TrpAlaGlyAlaProAlaSerAlaAla-----IleGlyThrAspIle 349

Db 1234 TTCAGGTCACCCGACCCGACCGGCGTCCAGCATTACAGCCGCTTACCGACCCG 1293
Qy 350 GlyLysLeuProSerMetMetProLeuGlnThrSerSerGlyAsnSerThrAlaArgAsn 369
Db 1294 AAGAAATACCGCGCGCGGACCATG-----ACCGGCGTGGCGCATGCCAACAC 1344
Qy 370 GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389
Db 1345 TCAGGCACC-----TACACTCAGCCGACGCTGACTGAC 1377
Qy 390 Val-----ThrIleThrSerGluGlyLysLeuGlnGly-----LysTyrSer 403
Db 1378 GTCGTCGATATCTACACCCCGCGGCGTCCAGGAGTTCAGACCGGCTCAAGTATTG 1437
Qy 404 AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423
Db 1438 ACCCGGTCAAACTGGTATTC-----GGT 1461
Qy 424 LeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
Db 1462 GCCACACAGCGGCGGACGAGCTTACAC---ATGGTTGATGCCAAGGGGCCACGATC 1518
Qy 444 Gly-----LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db 1519 GGTTCGCGGTCATCGTTCGCGTCAGGCGAACACCCCTGAACCTGAAGATCGCATGCTC 1578
Qy 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIle-----IleGln 477
Db 1579 GACTCCCGGTCGCTCCGTTG-----ATGGACACACCGTCATCCCGAACGTGCA 1629
Qy 478 ArgGlyPheGlnMetAsnSer 484
Db 1630 AAGACCTTCACCGTCGACACC 1650
RESULT 38
US-10-398-221-1330
; Sequence 1330, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398, 221
; CURRENT FILING DATE: 2003-03-27
; PRIOR FILING DATE: 2001-10-04
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1330
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Listeria monocytogenes-4B
US-10-398-221-1330
Alignment Scores:
Pred. No.: 2,16e-06 Length: 554
Score: 152.50 Matches: 57
Percent Similarity: 37.76% Conservative: 34
Best Local Similarity: 23.65% Mismatches: 81
Query Match: 5.90% Indels: 70
DB: 16 Gaps: 7
US-10-009-823A-1 (1-502) x US-10-398-221-1330 (1-554)
Qy 216 ValTyrAspGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla---Pro 234
Db 37 GTATATGATATGT-GCGGAAACATAAATCTCAAGTGAATGAAAGCTGCTACACCA 95
Qy 235 SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254

Db 96 GATCGGAGCGGTAATGTTCTTACGAATATGAATTCAAATG-----GACGGA 143
Qy 255 SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
Db 144 AAGACATTAATCTCTCCA-----GTAACAGAAACACTTAATTTACAAAT 185
Qy 275 SerAsnGlyLeuLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294
Db 186 GCGAGCGGCGAATTAACAAACCCAGACGCA----- 215
Qy 295 LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
Db 216 CTTAAAAATATCCAAATCAATTCACAGTAAACGGC----- 251
Qy 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
Db 252 -----AAACAAGTCAATATG 266
Qy 335 TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
Db 267 -----GGCTTAAACCTAAGTGGCTTAACCAAC 293
Qy 355 MetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThr 374
Db 294 TAC-----GGACAAACCAAGTA 311
Qy 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
Db 312 TTCTCACCACCTTCTGACGGTAAAGGGCGTGCAGCTGTAAGAGATTATGCGATTACCGAT 371
Qy 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
Db 372 TCTGGCTATATGCAATGAGTTACTCAGATGGTACAGTATCCAGTGGCCCACTTGCG 431
Qy 415 LeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAla 434
Db 432 GTGGCTACTTCTCCATGAGACGGCTTAGTCAAAATGGGAACGGCAATATGTTCCA 491
Qy 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
Db 492 GGATTATCTTCTGGCGATGCGATATACGGCTTGTCGCCAAATGCGCTGGCGGAATT 551
Qy 455 Ser 455
Db 552 AGC 554
RESULT 39
US-10-335-977-17
; Sequence 17, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...810
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-335-977-17

Alignment Scores:
Pred. No.: 4,74e-06 Length: 810
Score: 151.50 Matches: 58
Percent Similarity: 43.30% Conservative: 39
Best Local Similarity: 25.89% Mismatches: 90
Query Match: 5.86% Indels: 37
DB: 16 Gaps: 8

US-10-009-823A-1 (1-502) x US-10-335-977-17 (1-810)

Qy 6 PheleGlyAlaThrGly---MetLysThrHisSerThrGlyLeuGlyThrValSerAsn 24
Db 13 TATTATCGGCCCAACAGGCGAATGGGTACACAATTTAAACCGCTGGATTAAACCTTAAC 72
Qy 25 AsnLeuAlaAsnThrLeuGlyTyLysGlnGlnValValPheGlnAsp--- 43
Db 73 AATTAGCCCACTTAACACCAACGGCTTTAAAGAGACGATCGGATTAACGCGATTTT 132
Qy 44 -----LeuPheSerGlnAspLeuAlaIle 51
Db 133 TTAAGGCTTTACCAAGAATACCGAGACCACTGCCCTTAGAAGATCAACCAAGCGAGC 192
Qy 52 GlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerVal 71
Db 193 GCGAAGTATCTCAACCGCAACCTCAATCGTGTGCCCTATTCTATCAGAAATCTATACGGAT 252
Qy 72 ArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIle 91
Db 253 AGGAGTCTT-----GCGCGGTTGAAGGGACGAATAACCCCTAGATTTTGCCCTA 303
Qy 92 GlyGlyLysGly---PhePheGlnValThrLeuGluAspLysValHisThrArgAla 110
Db 304 ACAAGCCCTAACCTCTATTTTGGCAACAGACTAAATGAGGGCGCTCGCTTATACCAAGAC 363
Qy 111 GlyAsnPheArgPheThrGlnGlyPheLeuAsnAspProSerGlyPheThrLeuMet 130
Db 364 GGGCAITTCAGCGTTGATAAAGACGGCTTTTGTAAACCTTAATGTTTAAAGTGCTT 423
Qy 131 GlySerArgIleSerAsnAsn-----ProAsnIleLysLys 142
Db 424 ---TCAGCTCCCGTTTGAACGAAAAAGGAGGATCATGCTCATGCTCAACGCTGAAT 480
Qy 143 GluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAla 162
Db 481 GAAGTGATCAAAATGGTGGAAATCATTCTTTAGGAT-----NATGAAGCC 525
Qy 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182

Db 526 CAATTCAGCGCGCGCTTGTGCTTTAGTGGTAACTTAAATAATCTTAAAAA 585
Qy 183 SerGluAlaAsnProTyPheAlaLeuGluSerTrpLysGlyAsnGlyThrProPro 202
Db 586 ATAGGGCAAAACCTTTAT-----ACCTATCAGGGCGAAGGCGTTCAATCA 630
Qy 203 IleSerThrSer 206
Db 631 GTCTCTGACTCT 642

RESULT 40

US-10-282-122A-28180
Sequence 28180, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 28180
LENGTH: 6615
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-28180

Alignment Scores:
Pred. No.: 9,59e-05 Length: 6615
Score: 151.50 Matches: 119
Percent Similarity: 31.12% Conservative: 64
Best Local Similarity: 20.24% Mismatches: 220
Query Match: 5.86% Indels: 185
DB: 16 Gaps: 24

US-10-009-823A-1 (1-502) x US-10-282-122A-28180 (1-6615)

Qy 3 GlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22

Db 4372 GGCAATATCGGCATCGGCTCACCGC-----ACCGGTGAGATCGGGTTCGGGAGTTC 4425
 QY 23 SerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGln 42
 Db 4426 AACTCGGCAGCCACCAATCGGCTTGTTCACCTCGGTGACGGAACAGTAGATTCCTC 4485
 QY 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGly----- 55
 Db 4486 AACTCGGCAGCCGCAACGCTGGGCTCGGAACACACCGGCACCGCAAACTTCGGCATCGCA 4545
 QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 Db 4546 AACTCGGCAGCTTCAACACCGGCTCGGG-----AACCGGCAGCACCAACAGC----- 4596
 QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys--- 94
 Db 4597 -----GGCTGTTCACCGCGGCAAGTCACACCGGCTCGGCAACACACCGGAGCATC 4650
 QY 95 -----GlyPheGlnVal----- 99
 Db 4651 AACACCGGAGCATCAACACCGGAGCTTCAACACTGGGAGGACCAACATACCGGAGTTC 4710
 QY 100 ThrLeuGluAsp-----LysValHisTyrThr 108
 Db 4711 AACCTCGGCGATCACACACCGGCTGTCGCAACACCGGCTGACTACACACCGGCTACTTC 4770
 QY 109 ArgAlaGlyAsnPheArgPhe----- 115
 Db 4771 AACCGGCTGACTACACACCGGCTGTGGCAACACCGGCAACGTCACACACCGGCGTTC 4830
 QY 116 -----ThrGlnAspGlyPhe----- 120
 Db 4831 ATCTCGGCAATTACAGCAACGGGTTCCTCTGGGAGGTGACTTACCAGGGTGTATTGGC 4890
 QY 121 -----LeuAsnAspProSer 125
 Db 4891 CTTTCCACACGATCACCATTCCGGAATCCCTACCGCTAGCTAGTGTTCATTC 4950
 QY 126 GlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsn----- 139
 Db 4951 GACATACCCATCACCGGCACCGCTGTCGCCACCCACCGCAACAGTTTCACCATTCGCGT 5010
 QY 140 -----IleLysLysGluThrLeuGluPro 147
 Db 5011 TTCCAGATACGATCTGTGCTGCTCGGCTGCTGTTCACGAGATGATCGGCCCC 5070
 QY 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
 Db 5071 ATCAGCATCATGTCAATCAAGTCATCGCCATCGCCATTCGCCCATTCAGCAAAACCATCAGC 5130
 QY 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
 Db 5131 ATGTTGGCACCGGCGCTTCGGCCGATCCCATCGGCATCAGCATCGGTGTACCCCG 5190
 QY 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsn 207
 Db 5191 GGTTC-----GGCAACTCGACC----- 5208
 QY 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnValAsnSerHisAspIleThr 227
 Db 5209 -----ACCGGCCCTGCTCGGGTTCTTCCACACCGGCGGCCCATGTATCGGC 5259
 QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
 Db 5260 TTCGGAACTTCGGGCGCGCAACATGTGGGCTCGCGGAATTC----- 5304
 QY 248 MetAsnProSerGluAspGlySerAlaAlaSerGly---ThrAspSerAlaGlyLeuLeu 266
 Db 5305 -----GGCGCTGGCAATTCGGGCTTCTTTAACCGCGGCGGCTTGGC 5346
 QY 267 MetSerGlyThrMetThrPheSerSer---AsnGlyGluLeuLysAsnMetThrAlaPhe 285
 Db 5347 AATTCGGGCTACTGAATTCGGGCGCTGAGTCGGGTCTGGGGAACCTGGGCAACACC 5406

QY 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
 Db 5407 ATCTCGGGGTCTACAAACAGAGCAGCTCGGACCTCGGACGCCGCC-----TTCGGCTCG 5463
 QY 306 GlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln-----ProLeu 322
 Db 5464 GGATCGCAAAACATCGGCGCCAACTCGGCGGCTGTCTCTCGACAAACACCGCAACCTG 5523
 QY 323 ThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342
 Db 5524 ACGTGAATTCGGCTCGCAACACCGGCGGCTCAACCGCGGC----- 5568
 QY 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362
 Db 5569 -----ATCGGGAACCTCGGCAACCTCGGCGACCTCAACTTCGGCGGGTCAACATC 5610
 QY 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyr----- 377
 Db 5611 GGCAGCTCAACACCTGGGATCGGCAACCTCGGCGACCTCAACTTCGGCGGGTCAACATC 5670
 QY 378 -----SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
 Db 5671 GCGGTAAACATCGGCATCGCCCAACACCGGATCTTCGATATCGGCTTGGCAACCTG 5730
 QY 396 GlyLysLeuGlnGlyLysTyrSerAsn-----SerGlnValValAspPhe----- 410
 Db 5731 GGCAGCTCAACATCGGCTTGGCAAACTCTGGGCGACGACCAACCTTGGCAACGCC 5790
 QY 411 -----TyrAsnIleProLeuAlaArgPheThrSerGluAsp----- 422
 Db 5791 GGCAGCTCAACATCGGCTTGGCAAACTCTGGGCGACGACCAACCTTGGCAACGCC 5850
 QY 423 -----GlyLeuArgGluGluGlyAsnAsnHisTyrSer----- 433
 Db 5851 GGCAGCTCAACATCGGCTTGGCAAACTCTGGGCGACGACCAACCTTGGCAACGCC 5910
 QY 434 -----AlaThrLeuAspSerGlyGlyProGluPheGlyLeu--- 445
 Db 5911 GGCAACGGCCAGATCGGATCGGCGACCTCAACTCGGCGGACCAACATCGGCTGTTTC 5970
 QY 446 ---ProGlyThrSerAsnTyrGly 452
 Db 5971 AACTCGGCGAGGAAACATCGGG 5994

Search completed: October 31, 2004, 03:22:14
 Job time : 2291 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2004, 03:11:41 ; Search time 4223 Seconds

(without alignments)
5621.461 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMGSLFIGNKTHSTGGLG.....NSKSVTTADTMQLKALELKR 502

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2/1/USFTO.spool/US1009823/runat_26102004.100129.1603/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPL=0 -ICOREXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdd -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1009823 @CGN 1.1 2527 @runat_26102004.100129.1603 -NCPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.btg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1235	47.8	300343	1	AE017310 Desulfovi
2	961	37.2	300029	1	AE017314 Desulfovi
3	693.5	26.8	110000	1	AE017180.04 Continuation (5 of
4	595.5	23.0	9612	1	AF122909 Treponema

C 5	595.5	23.0	123663	1	AE017255
C 6	579.5	22.4	349965	1	BX842655
C 7	567.5	21.9	2516	1	TEU04619
C 8	563	21.8	110000	1	CR522870_30
C 9	557.5	21.6	12559	1	AE013853
C 10	557.5	21.6	290155	1	AE017132
C 11	557.5	21.6	335050	1	AE017132
C 12	550	21.3	10029	1	AE011450
C 13	550	21.3	300594	1	AE017291
C 14	534.5	20.7	5230	1	TEU28219
C 15	534.5	20.7	26536	1	AE001244
C 16	533	20.6	313050	1	BX321857
C 17	528.5	20.4	10303	1	AE013102
C 18	527.5	20.4	25022	1	AE004539
C 19	509	19.7	34817	1	BBU43739
C 20	509	19.7	44380	1	AE001137
C 21	503	19.5	1499	1	BORFLGE
C 22	502	19.4	203050	1	AL646078
C 23	498	19.3	1320	1	STFLGR
C 24	498	19.3	20753	1	AE008751
C 25	498	19.3	301983	1	AE016840
C 26	498	19.3	306358	1	AE016759
C 27	498	19.1	9224	1	AE005317
C 28	493	19.1	222605	1	AE002555
C 29	493	19.1	110000	1	U000396_11
C 30	491.5	19.0	110000	1	AE001509
C 31	485	18.8	19549	2	AC095028
C 32	485	18.8	90250	2	AL591976
C 33	484.5	18.7	250050	1	AL591976
C 34	484.5	18.7	260050	1	AL596166
C 35	484.5	18.7	290242	1	AE017324
C 36	484.5	18.7	349980	6	AX417038
C 37	484.5	18.7	349980	6	AX417039
C 38	484.5	18.7	349980	6	AX417042
C 39	484.5	18.7	349980	6	AX641666
C 40	483.5	18.7	13374	1	AE015135
C 41	483.5	18.7	292504	1	AE016981
C 42	483	18.7	2550	1	HFU09549
C 43	482	18.6	110000	1	BX950851_19
C 44	482	18.6	346613	1	BX571661
C 45	481	18.6	8513	1	AE012299

ALIGNMENTS

RESULT 1	AE017310	300343 bp	DNA	linear	BCT 27-APR-2004
LOCUS	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough, section				
DEFINITION	2 of 12 of the complete genome.				
ACCESSION	AE017310	AE017285			
VERSION	AE017310.2	GI:45577667			
SOURCE	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough				
ORGANISM	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough				
REFERENCE	Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;				
AUTHORS	Desulfovibrionaceae; Desulfovibrio.				
	1 (bases 1 to 300343)				
	Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L.,				
	Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B.,				
	Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Bodson, R.J., Durkin, A.S.,				
	Nadapu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H.,				
	Selengut, J., Peterson, J.D., David, T.M., Zafar, N., Zhou, L.,				
	Radhakrishnan, S., Dimitrov, G., Hance, M., Tran, K., Khouri, H.M., Gill, J.,				
	Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and				
	Fraser, C.M.				
TITLE	The genome sequence of the anaerobic, sulfate-reducing bacterium				
JOURNAL	Desulfovibrio vulgaris Hildenborough				
PUBMED	Nat. Biotechnol. (2004) In press				
REFERENCE	15077118				
AUTHORS	2 (bases 1 to 300343)				
	Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L.,				
	Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B.,				

Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., Davidsen, T.M., Zafar, N., Zhou, L., Radune, D., Dmitrov, S., Hance, M., Tran, K., Khouri, H.M., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.

Direct Submission

Submitted (18-MAR-2004) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
On Apr 27, 2004 this sequence version replaced gi:46448084.

FEATURES

source

Location/Qualifiers

1..300343

/organism="Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough"

/mol_type="genomic DNA"

/strain="Hildenborough"

/sub_species="vulgaris"

/db_xref="taxon:882"

753..1217

/locus_tag="DVU0257"

753..1217

/locus_tag="DVU0257"

/note="identified by match to protein family HMM PF00583"

/codon_start=1

/transl_table=11

/product="acetyltransferase, GNAT family"

/protein_id="AAS94740.1"

/db_xref="GI:46448085"

/translation="MQSIPCKVPLTQGHYSAYEAVESVSTHAFLEDDLTIRA
RLVSDLEPAVTLGVFVWEGSLAGPAGTASKEMLFIATPWQGRGLRLVESAIRHE
GVAFVDNEQPRALATYRNGFEVGRSGRGMGPYPLHFWLRHEDQK"

complement(1378..3351)

/locus_tag="DVU0258"

complement(1378..3351)

/locus_tag="DVU0258"

/note="identified by match to protein family HMM PF00072;

match to protein family HMM PF00312; match to protein

family HMM PF00785; match to protein family HMM PF02518;

match to protein family HMM TIGR00229"

/codon_start=1

/transl_table=11

/product="sensory box histidine kinase/response regulator"

/protein_id="AAS94741.1"

/db_xref="GI:46448086"

/translation="MSESLVDDERGIKVLGITLADAGYRVTTAAGEEAMRRFAE
ERPDIVTIDKPMGMSLELESIKRDSSEVIMLTGHDMELATSLKRDATDPT
KPINDDVLDALGRARITMRRLREYENLEDVAKQSAALVAERQALALQVMDG
IASGVOTSLMDQAGIFNELPCFVSNHGRYLEIVATNDLYKERLGHKVGQPSWEVYK
GREGSGNCPVWVETGEGORNEILIDKHGOEIPVIVHTAPILNDGEIEMVLEIS
VDVSKVRLQDELRITREYRQLEDEAPCVIADRSKLVADANREHDFGEPAAEH
CHDIFARLACACGCPVERSDGKPHQYETVTTGDSQANVLVWTSPLRDAAGNIT
EWMESDITQVQLQRLQSLGLLSTAGHKGILTLDDGVYRLGSLGRDNDEAR
VRQSPEDIRLTLRNVLDLILYAKKRDLEWVVVAERFATETAELEGGVANGV
GQLEIAEGSTFEADAGALSALVALLNENAVACAARSPEHSVTFVSGVDANHE
FSVRDGVGDRETRKELFTLFFSSKSGAGTIGLGFVANQVQQHGHIAVASPEGHG
TTFVSLRPLPVARCPAQNGQPEGDEADDN"

complement(3612..4001)

/locus_tag="DVU0259"

complement(3612..4001)

/locus_tag="DVU0259"

/note="identified by similarity to SP:P33394; match to

protein family HMM PF00072"

/codon_start=1

/transl_table=11

/product="DNA-binding response regulator"

/protein_id="AAS94742.1"

/db_xref="GI:46448087"

/translation="MAYKILVDDDDPYIVKVMVDLSNDNGYATCSASDGEALVLR
EKPDVLTDLMEVNEGPRFYKRWTFREDFRNVPIVISGLSGIHLAINVASLKKP
FDNKLLEIRIQLENEPGRSAGEH"

complement(4019..4420)

/locus_tag="DVU0260"

complement(4019..4420)

/translation="MAYKILVDDDDPYIVKVMVDLSNDNGYATCSASDGEALVLR
EKPDVLTDLMEVNEGPRFYKRWTFREDFRNVPIVISGLSGIHLAINVASLKKP
FDNKLLEIRIQLENEPGRSAGEH"

complement(4019..4420)

/locus_tag="DVU0260"

complement(4019..4420)

/locus_tag="DVU0260"

complement(4019..4420)

/locus_tag="DVU0260"

complement(4019..4420)

/locus_tag="DVU0260"

complement(4019..4420)

/locus_tag="DVU0260"

/note="identified by match to protein family HMM PF00072"

/codon_start=1

/transl_table=11

/product="response regulator"

/protein_id="AAS94743.1"

/db_xref="GI:46448088"

/translation="NRVLIVDEMGMRFLRLTLPETGCGHEAVVARGOEGLARAREIR
PLHVLVDMVPMGGGLVYRALRDADLAATPVLMSGVKGTFAHALELYDAAAGTL
PPEAYVEKPPRPDEILALAEPLTGTCTGSCG"

complement(4440..5366)

/locus_tag="DVU0261"

complement(4440..5366)

/locus_tag="DVU0261"

/note="identified by match to protein family HMM PF00582"

/codon_start=1

/transl_table=11

/product="universal stress protein family"

/protein_id="AAS94744.1"

/db_xref="GI:46448089"

/translation="MFKIIFATSNASPADNAAKVAFLAPBNKARLYVHVLCVPSR
GFSQVRLIRTGESEHHDADYRDVLEMKOTYSYQLETYSQVPMVQVAVGVPHEI
LRFARQEGADLIVMGANTREDPGASRVRSIGNTMQVARSAPCVLIINRPTCTW
NLFNIVSTDFSKAASAFALKTAQIGAKLYLHACDLAANPMGVFAGQVEIER
RIEKAALMQERYYSKMGDFENYQVDIWEGTYPYVILLKYARARQADLIUVAHTKEG
VEEABIGSTVQVWLRSPVAVSKPEKAVV"

complement(5370..5924)

/locus_tag="DVU0262"

complement(5370..5924)

/locus_tag="DVU0262"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAS94745.1"

/db_xref="GI:46448090"

/translation="MDDIRLGNDCGILREMGTTVADPLACLRNVTLDEGYSURGYFA
LLGVHPLVMSPPFLPALRENDGACGCLTESFERELGRTVELGFDDAARVEVY
HTLRGITGEAGSPARSEIRFVPLGCLLDMPLHGRARHVLFGAARVLECDTTFRL
FDIIEGLAWELGPOGSGQCSLGR"

complement(5946..6326)

/locus_tag="DVU0263"

complement(5946..6326)

/locus_tag="DVU0263"

/note="identified by similarity to SP:P94690"

/codon_start=1

/transl_table=11

/product="acidic cytochrome c3"

/protein_id="AAS94746.1"

/db_xref="GI:46448091"

/translation="MFRIRIGTVMALFATALLAVAFARTDTPRLEPAALKPQORPAA
VFDHDAENKALENCVCHHDAKDKIVPEGSEGTPCADCHAKATKGTILNVAVH
RQCNCHKTSGKGTSCGCHVRD"

complement(6352..7671)

/locus_tag="DVU0264"

complement(6352..7671)

/locus_tag="DVU0264"

complement(6352..7671)

/locus_tag="DVU0264"

/note="identified by similarity to SP:P33393; match to
protein family HMM PF00037; match to protein family HMM
PF02754"

/codon_start=1

/transl_table=11

/product="ferredoxin, 4Fe-4S, putative"

/protein_id="AAS94747.1"

/db_xref="GI:46448092"

/translation="MGITERIEDAGLMRGIAALTPERIEKVVKVAITEGEGARLKVY
YETCMRGCLGACHYICISHDGPDYSPVGMKEQTMWRLREGGRVSPDIYDMAQIA
YTCNCRRCIHYCPIDGTGYVSLVRIICHLGVTTPRYMQTAISHAATFNQMWVK
EDWDVSLIWOEDEARSEIPTMRIPLEVAGADILYSVIAPLQPKFRTQLIYQAAVIMDQ
AGYSWTMPATPCGNDNDMAVFGSDSVMGELIKPAHFEALQRLVKELIWMGCGHAFRS
IYDVGRLWLRKMHVPVHVSVEFWELFDQKIKLAKRFEPEVTHIDFCNIVRGRL
MDKUREVVHALCENNVEMTPEHNFNCCCAGGGVINCPCGPFKGRVVEGNRIKAEQKA

```

TGKTVIAPCHNCHGLEPDIHHYKLGIMHTKFLGDIIVDCMEKPGAV"
complement (7691..8350)
/locus tag="DVU0265"
CDS
complement (7691..8350)
/locus tag="DVU0265"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAS94748.1"
/db_xref="GI:46448093"
/translation="MOAFNLLIPGAMTAFVFLGGCAWRLSTMGALARRKDMVAVA
YMSKSLMSIMHWTPPTLGRNENPAMTVAEPHISIVILALPAPHAATMWDYAF
GVGWALPRAVDLITIVLAACGPAPRRLSDPVSVFTRPMDNLVLGLAARVLTG
ILAKYVMTQNLVAFTHLVUSGAMLVALPFLRLSHAVSFVTRVIGSEFGVRECKD
W"
complement (8359..9585)
/locus tag="DVU0266"
CDS
complement (8359..9585)
/locus tag="DVU0266"
/notes="identified by Glimmer2; putative"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS94749.1"
/db_xref="GI:46448094"
/translation="MKRKYIASVAPLEHLMQBEAQPAPDGSAPAAVVCVEGAFSL
RVNDIVMDTFEKAMTPRTPDRLTALVQIDGWTFLAVDGPWEETGFWMDTHFGS
Alignment Scores:
Pred. No.: 1-566-68 Length: 300343
Score: 1235.00 Matches: 253
Percent Similarity: 64.12% Conservative: 74
Best Local Similarity: 49.61% Mismatches: 163
Query Match: 47.76% Indels: 20
DB: 1 Gaps: 5

US-10-009-823A-1 (1-502) x AE017310 (1-300343)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
DB 51372 ATGATGACGAGCTCTACCGCGCGCCAGCGGATGAGCAGCGAGCCGCAAGATGATGAA 51431
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
DB 51432 GTGGTGGTAACAACCTCGCAACGTGACACCATCGGCTTCAAGCAGTCCATGATGCTG 51491
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
DB 51492 TATCAGGACCTGATGAGCCAGACGGAACCCACCGGCAGTGCCTATCTCTCCGGCATATCG 51551
QY 61 GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPhe 80
DB 51552 CAGTGGGACTCGGTGCGCGGGTGGCGGACGTGCGGTGCGTGGCGGGGCGCTTG 51611
QY 81 GluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnValThr 100
DB 51612 CTTGCGGGTGGACACACCGACTTCGCCATATCGGGCAAGAGCTTCTTTCAGGTGACC 51671
QY 101 LeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPhe 120
DB 51672 AGCGGGCAGACACCCACTACCCCGCGCGAGCACTTCGGCTTCAACAGACGCGGCAA 51731
QY 121 LeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIle 140
DB 51732 CTGCTGACCCCAACAGGGTTCACCTGATGGGCGACGCCATCACCGT----- 51779
QY 141 LysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer 160
DB 51780 -----GAAACCGAAGCCCGCTGGCAGCC-----GTGACGCTGGTGAAGAT 51821
QY 161 -----ProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173

```

RESULT 2
AE017314/c
LOCUS

```

51822 CGGACCGGGCGGTCTTTCCATCGCCGCCCAAGCGACAGGGGCGATGACTCTGTCTTCAAC 51881
174 LeuGlyAspSerThrAspLysThrGlnSerGlnAlaAsnProTyrPheAlaLeuLeuGlu 193
51882 CTCGGTTTCGACAAGGATGCCAGCAGCAGCAGCCCAATCGTCTTTCGGATCTCGAG 51941
194 SerTyrLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
51942 GCATGGGACCGCACCGCTACCGCCCTTGGCGACATCGGCGAGCTACACCCAGACC 52001
214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
52002 ATGCGCGTCTATGACGTGGCGGACACCCACCACTCACCATCGCTTCGACAATGCC 52061
234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
52062 ACCACCGCGCGCGCATCGTGGAGTTCTCGTCCACCATCGCTCCCGCGGAGGAC 52121
254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
52122 GCGGTGCGCGCGGTAGCGCGGTGCGGGGTGCTCATGTCGGGAGCGTTCGAATTC 52181
274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
52182 GGGTCGTGAGGCGAGTTCAGGACATCATGGCCCTTCAAGCCATCAGGCGGACCCCAAG 52241
294 AspLeuAsnAlaTatGlnProAlaProLeu--ValAsnGlyLeuProGlnPheSerAla 312
52242 GACCTCGCCAACTGGTGCCTCCGCGCAGCTGACCCCTCTGGCAGCGCGCAGTTCAAGCC 52301
313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332
52302 ACCTTGGCGAAGCGCGGCGCAGACAGTGGCCCTCGACCTTGGCATCACCACC----- 52355
333 AsnMetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352
52356 ACCCATGGAACAATGACCCGCTTCGGCGCGGGGTGGCGCTCGCAGCCGCGCTCTT 52415
353 ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 372
52416 GCGGGGCGCCACCGCCAACTGGCTGCCACGTCGACCCAGCTTACAGGGAAGTTTCG 52475
373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392
52476 TCGTCGACCTACATGAAGCAGCAGCGCTATGCGGAAGCGCTACTCATGAACCTTGAGTG 52535
393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412
52536 GCAACGATGGCATCGTCAGCGGCAAGTACAGCAACCGTCACTCGAGTCGCTCTCCGC 52595
413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyr 432
52596 GTGCCCATCTCCGCTTCCACCGAGGACGCGGTTCGGAAGCGGCGATGAGGCGCATGAC 52655
433 SerAlaThrLeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
52656 TCCGCGCACCGAGTCTGGTGGCGGACGGAAGCGGCGGCGGACAGAGAGATTCAGCGC 52715
453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
52716 AAGCTCATGGCAAGTCTTTCAGCAGTCCAACTGGACATGGCCCGCGAGATGGTGAAC 52775
473 MetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAsnThr 492
52776 ATGATCGTACCCAGCGCGGATTCAGACCAACAGAGAGGTGGTCCACCGCGGAGACC 52835
493 MetLeuGlnLysAlaLeuGluLeuLysArg 502
52836 ATGATTCAGAGGCGCTGGAATCAAGCGC 52865

```

AE017314 300029 bp DNA linear BCT 27-APR-2004

DEFINITION	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough, section 6 of 12 of the complete genome.
ACCESSION	AE017314
VERSION	AE017285
KEYWORDS	AE017314.1 GI:46449255
SOURCE	
ORGANISM	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough
BACTERIA	Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacteriales; Desulfobacteriaceae; Desulfobacterium
REFERENCE	1 (bases 1 to 300029)
AUTHORS	Heidelberg, J.F., Seshadri, R., Havenan, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., David, T.M., Zafar, N., Zhou, L., Radu, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H.M., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.
TITLE	The genome sequence of the anaerobic, sulfate-reducing bacterium Desulfovibrio vulgaris Hildenborough
JOURNAL	Nat. Biotechnol. (2004) In press
REFERENCE	15077118
AUTHORS	2 (bases 1 to 300029)
Heidelberg, J.F., Seshadri, R., Havenan, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D.E., Haft, D.H., Selengut, J., Peterson, J.D., David, T.M., Zafar, N., Zhou, L., Radu, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H.M., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G. and Fraser, C.M.	
TITLE	Direct Submission
JOURNAL	Submitted (18-MAR-2004) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1..300029
	/organism="Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough"
	/mol_type="genomic DNA"
	/strain="Hildenborough"
	/sub_species="vulgaris"
	/db_xref="taxon:882"
	complement(170..1270)
gene	/locus_tag="DVU1429"
CDS	complement(170..1270)
	/locus_tag="DVU1429"
	/note="identified by match to protein family HMM PF06071; match to protein family HMM TIGR00092"
	/codon_start=1
	/transl_table=11
	/product="GTP-binding protein"
	/protein_id="AAS95907.1"
	/db_xref="GI:46449256"
	/translation="MALSIGVLPNVKSTLFNALTKAQNAPAAAPYPTIEPNKAT VAVPDLALATIVKPVQVHATVPIDIAIGVRSKSGEGNLFNLANEACRAIL EVKRFDEMTNVDGADIDRLDIEETIELLADTQSAEKPERLQKMAKGDKAKA AEEVAPLLRHMAKGPASTFEARESDAFOLAREGLITAKKIYYCANVDITGEE NAPHVRLVRELAASRGADVVKCAKIBEELOGLSDEQASMLSGYIDEGSLVSIITG YHILGASVITRGEKVRAMTIKQGNKAPGAGVHIDFERGIRAEVIAVDYVKHR SEACRSAGVIRVEGVEYVQMDGDVHFLFNV"
	complement(1684..4290)
gene	/locus_tag="DVU1430"
CDS	complement(1684..4290)
	/locus_tag="DVU1430"
	/note="identified by match to protein family HMM PF05193; match to protein family HMM PF05193"
	/codon_start=1
	/transl_table=11
	/product="peptidase, M16 family"
	/protein_id="AAS95908.1"
	/db_xref="GI:46449257"
	/translation="MRIALFTCTIIGVMLATSVQAAQPADITRLSNGITVLVKDDR FPLASRLYHVASAFETPEAGISHLHEWFKGVKPKGVARDVESVGYLNA

gene	TSFDVTVLTDPSTQWKLGMVDLDRWAFALDPAELESSEKDVVIAELQREGSDPS RIFQSOGATLKGTYVERPIGTETIRATTATMTRAVIRKHYQSMMLTVGVNDP AEVRAEABRLPGGLTNDQNTTPPAADARAFAGPVVNVPHGPMKPYLVGLVAPVGL KALQAQDMLSLGGPTALLRTFKYERQLQVDDISVANSYSEFVGMVLYIATLDA DKVETFWKELTLMLAGDKADFTQEFERARLNEDGLYRSKETVAGLSKRGFFQFF SDQFGZANYLQAVRDERPQLEALIRDLWRPERLTVSVLLPECTKAPALAAATKSTW SKGKAABEGNAAGVGTETVDLGGKRVLLIPDTILPYTALDMVFGSGNALDQNRQ GLAATASVLTGTLKHDATPLEAFQSDRAASLGASAGRRFTLSLREPFRQDQMG LLEHVLTALAPDEVAERENOVASIRAREDOPLGAERHLTPFLPGHSQVGYHLG QPTEVFETRDDVAKAFWARQAAQPMVMSVAGSDREAVLPKSLPAPSKVPYSLDAP AWTEPKALDULPERNOAHLLLVFTVGLAHKUTPALELLQSVLUGSGSLFRDMRK QGLGYTAVMWNQSLDLAGFVWLYGTGPEGLQAEAGFRKVIDQLHATALPDELRKG KNQMGDYREHQRLGSRSEAAMLTSQGYPLLFNREVIDKAEKLAFLPSDLERVATYL DMNKVRVVKVLP"
CDS	4484..6679
	/locus_tag="DVU1431"
	4484..6679
	/locus_tag="DVU1431"
	/note="identified by match to protein family HMM PF01627; match to protein family HMM PF01740"
	/codon_start=1
	/transl_table=11
	/product="hpt domain protein/STAS domain protein"
	/protein_id="AAS95909.1"
	/db_xref="GI:46449258"
	/translation="MTDOEKRTQLERFELGEALRVLDADAGVDLLELVGGDGEHVAVL GRLSRITQWRGAGLGNAMHEALVLEAVLDLARHEKPLDITVITAIERGMDF AARFKMRDSQDAGTAEVTOAQLQGVVQPADARCI SERHVVRFQDVDAFT MGGHDLARASREGLRVFVLAEDVHDLQKQVSPGLLVFLQKSGVDVDFGVGG GGGNVGLFSLIDALLGAVFOIDASKRFAVDTDVARGEAAARERDLAPAGKSA MPDPSRSIELLLQLYDEAMKRLREQSMQDPEPYDMNSDTDEKPAQONAGFEV VMEDTSDGQEAEMSLGSLTSDVADPVYQAAGAEAEYATDAPSSOMAGDAMPPLDA PGSYFFDDAPGVGLITDTPHPVADITLGEASGLAGDDADVDDDAASLULLEAEF DAPLLDADAYAVEPHRIGLAEGTLDGAASHLSGLSIADSLDLPDDADPPDDSH LDPLSRGKPGGKGLMLDGFDDAAGDEAPVDTPEAPAPFAAGMGLASLVDP DVMAVGAAGGLPPHDEAIAEAVPAGTSPATSVSPSPVSGTGAAGAEASGGFD VRESGSGILTLGSELTIERSLFRDALLEMFDYQNVRLDLSGVTEVDLTFQLLRA AVVTAQRGGTLAGGVVAPVSETAHRAGLDAQAKRAGFEAVLGMA"
gene	complement(6906..9500)
CDS	/locus_tag="DVU1432"
	complement(6906..9500)
	/locus_tag="DVU1432"
	/note="identified by match to protein family HMM PF04055"
	/codon_start=1
	/transl_table=11
	/product="radical SAM domain protein"
	/protein_id="AAS95910.1"
	/db_xref="GI:46449259"
	/translation="MRELLSLPKPSRYTGIEGSHKRRVEDVAIHCAALFAPDMYEVG MSYLGKILYLINDREDIWAERVFTPCREAGAILRSHNAPLSTLSDTTPLAALHLVG PSITHELCTVILYMLDLAGILPLRSADRGDDLAAPTIIMAGGGCTLAEEAPWFLDLM VLGESEVOVEIIDLAKABSEGMSPTYLEAASSIPGVVPAFFADAGATPAPLR QGDKVVRRIADMDNARYPAQQLVPFGAVHNLALIEIGCGTRGCGPCOAGMTYRPA RERSVMDALLEDCLARTGTDVSLSTGDSFSAKLTLFQTVDRCAEQVSLSL RSRVSGDDDDIMRMAGIRGTATLAPAGSQQRDVRVINKVTEEGMLHVRKLFHFG WQQVYLMGLPTTETEDLADVLCKRVDAAGRVRLQVTAASVPVFKPFPF QWEAKIIELETRARVGLRDARREKGLRWHPEPMSFLEIGFSRGRDLAPVESA REKALIFSSNMGEFDLAPLWEMNHEGLSPDEYTCARDIDAPLAWHLDSGVREFL RERKALIKTTDPCRYSGCVCCTKAGPSRLDPLGVAAYANLNFQDQOSH EPTLDDNGQVVRHEAFNRPKLSGDLVHKASQVRYWTKEDMAVFLSQLEOTIFERA LRRAGLPAPSQGFHPLPLLSFGALPVGASRAEWNTIILRRFTAFVROQLGRL RKNRMTVFVEEQLPQKHPQVVEFTRYLPAPEWQPTFDIWARFTFAACEAFEWDR LTKGQRTGDIRPLFRDITTFDDGSLTLRLDWTDFYMSPLARAVPFAAINELAMTK LEQHFEPA"
gene	9499..9690
CDS	/locus_tag="DVU1433"
	9499..9690
	/locus_tag="DVU1433"
	/note="identified by Glimmer2; putative"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="hypothetical protein"

```

/protein_id="AAS95911.1"
/db_xref="GI:46449260"
/translacion="MSVLQVAVVATKERVASGHPQKQKAPVSGGLRHQRQRHRVAGAI
GATGSGISVSGAAGROQLWL"
complement(9736..10101)
locus_tag="DVU1434"
locus_tag="DVU1434"
complement(9736..10101)
locus_tag="DVU1434"
note="identified by Glimmer2; putative"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS95912.1"
/db_xref="GI:46449261"
/translacion="MAYDIIRLVKLVNGELVIGKFDADAYKDVAALQITPTQOGVM
MLMPYGVFFGQPGVIERHGVYVEYKSPDDLKTKMEACSNLTLTSGSLGGVGLG
LGKFGMGKYSDLSSLKK"
complement(10164..11348)
locus_tag="DVU1435"
locus_tag="DVU1435"
complement(10164..11348)
locus_tag="DVU1435"
note="identified by similarity to OMNI.NTLC3PA01784"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAS95913.1"
/db_xref="GI:46449262"
/translacion="MSTSTSPKALPTAFVALLFMLNPTARSALSPLLVDLERLD
AVGHNQATGLLVQAQFSQAVCGVCGELLARIISPRMVALSITLGLCLIGMSLVLT
CTARLVGLFSGFMAGLVFPAGMATLSIVSPRDGKAVAVHELAPNTSTFLPLPRAE
ALRHMWRGVSPVIGSLCGLALGLFALFGRGGRTLAAPPYAGCRALRTPTATWVIV
LFVSCVAFGEFTFVSLVPLVLTSHFTTETANNQVLSLRLLCPFAAVAGWGFADRAQ
GGITKGVILLHGAALSMLSEPVVVFIMGSAQAFVTFSPFALFKFAQSFPAARQ
RLSLITWPTVACLITGTAAPFLGCGQHAGFGAGFMALGGISVLTLLTLRALPHALG
RG"
complement(11485..11580)
locus_tag="DVU1436"
complement(11485..11580)
locus_tag="DVU1436"
note="identified by Glimmer2; putative"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS95914.1"
/db_xref="GI:46449263"
/translacion="WTAEGRPRAKAPGWAAGMTAGIPADWVIDTT"
11769..11867
locus_tag="DVU1437"
locus_tag="DVU1437"
complement(11769..11867)
locus_tag="DVU1437"
note="identified by Glimmer2; putative"

```

Alignment Scores:

Alignment scores:		
Pred. No.:	9,966-51	Length:
Score:	961.00	Matches:
Percent Similarity:	53.8%	Conservative:
Best Local Similarity:	38.9%	Mismatches:
Query Match:	37.1%	Indels:
DB:	1	Gaps:
		300029

US-10-009-823A-1 (1-502) X AE017314 (1-300029)

Qy 1 MetMetGlySerLeuPheIleClyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Dh 17948 CTTACCGCAAGCATATGGACACGGCGCTTTCAGGATCTGCCCCAGGCAGAGATGAAC 17889

21 ThrValSerAsnAsnTleAlaAsnAlaAsnThrIleGlvTyrIvsGlnGlnValVal 40

17888 GTCTCGGCACCAACATCGCCAAAGTCAACACCGTCGGCTTCAAAGGCAGCCGTATGGAC 17829

41 pheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60

Db	17828	TTCCAGAGACTTCCTCAATCAGGAC-----ACCTACAGCGCAGCGGGCGTGACG	17781
Qy	61	GlnAlaGlyMetGlyYalaGlnValGlySerValArgThrIlePheThrGlnGlyYalaPhe	80
Db	17780	CAGGTGGGTCCGGGTGAGCATCGTGCATCTTCGGGGACTACAGCAGCGGGGCGCTTC	17721
Qy	81	GluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnValThr	100
Db	17720	CAGACCAACCAAGAAATCCACCACTTCCCAATTCAGGAGCGGGGTTCTTCTCGGTGAAG	17661
Qy	101	LeuGlu-----AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAsp	118
Db	17660	CCCAAGGGAACGAAGACACCTACTACACCGCGCGGGCACTTCAGGTTCGATCCCGAC	17601
Qy	119	GlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnPro	138
Db	17600	GGCTACTCTGTAGACCGCATCGCTCACTCGTCTGCAGGGGTGGCCATCGAGCGCAGCAG	17541
Qy	139	Asn-----IleLysLysGlu	143
Db	17540	AATTCGCTCACCTGCTCCGCTGTACCGCCACAGCACCGTCCAAATCAAGGGTTCC	17481
Qy	144	ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLys	163
Db	17480	GGTGTG---CCTGTGACATCAAGCTC---GACGGTTTCACCGCGGAACCGCAGCACACC	17427
Qy	164	ThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSer	183
Db	17426	CAGAAATCATCGCTCAATGTCAACTCTGCAGCTCCATCCGGGTAAACGAAAGACGACAGC	17367
Qy	184	GluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGly-----AsnGly	199
Db	17366	ACCACGAACCGGTATTTCTCGCTCTTCGAGACATCGAAGCGGCCAGAACCCCTCACGGGC	17307
Qy	200	Thr---ProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAsp	218
Db	17306	ACACAGCTTCGCTCCGCCAGTCGGCGTATCGCTATCATGTCACCATCAAGTCTATGAC	17247
Qy	219	GlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla---ProSerSer---	236
Db	17246	GAAAGCGGTACCGCCACACCTCACCTCACCTTACTTTCGACCGAGTCGACCCCGATAGCGTG	17187
Qy	237	-----ThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGlu	252
Db	17186	ACCAAGAACCCCAATGGTCGCAACAAATGGGAATACATCGTCACCATGGACCCGCGCCGAG	17127
Qy	253	Asp-----GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet	267
Db	17126	GACAAGCGGTATCATCCCGGTACGGCATGAACACCATCGGGCGGGCGCTGCTTAATG	17067
Qy	268	SerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrPro	287
Db	17066	ACGGGTACACTCAGTTTCGACACAAACGGGGCAGCTTGTGACACGAGACCGCTTCACCCC	17007
Qy	288	ThrGlySerAlaThr-----	292
Db	17006	TTCCGTCAGTACAGGACACCAACCCCGTGGAAAGAACTACGCCCCAACCCCGCGCCG	16947
Qy	293	-----LysAspLeuAsnAlaTrpGlnProAlaProLeu---ValAsnGly	306
Db	16946	CCTGTACACAGCGCTCCGACCTCGTCACTGGCAGCTTACGAGATGTCCTCCACACGA	16888
Qy	307	LeuProGlnPheSerAlaAsnPheValGly-----	316
Db	16886	CTTCCCATGATGGTGGCAACTTCTCGGCGCTTACCGACTCGAGGCGTGTAGGTTCCTCC	16827
Qy	317	---AlaGlyIleGlnProIleThrLeuAspPheGlyIleLysSerGlnGln-----Asn	333
Db	16826	ACGGCACAGAACTTCTCTCATGGAACTTCGACCTTCGGCTTGGCTTCGACCAACCGGACACG	16767
Qy	334	MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro	353

Db 16766 CCGTGGACATCACCCCC---AACCGCCGCCCATCGGTACGGAGCGCTCGCTGTCGCG 16710
Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaAaGAsnGlySerSerSer 373
Db 16709 GGTCTGACCTTTCGCGAGCCGCCAGCCAGTCACAGACAGCTACGCGGCTTCGTATCG 16650
Qy 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 16649 ACACAGTTTCCAGAACGAGCGGTACACGTTCCGCTTCTTGCAAGACATCATCTGTCGAC 16590
Qy 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 16589 CGTGACGCGATCATCGAGGGGAATATCCAACGGTGTGACTCTCGACCTCTATCAGTGC 16530
Qy 414 ProLeuAlaAaPheThrSerGlnAspGlyLysLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 16529 ACCCTCGTGAGCTTTCACAGCAAGCAGAACCTGCGCGGAGAGGGCGGCAACCTCTTCAGC 16470
Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn----- 450
Db 16469 GGCACACAGATTTCGGGACCTT-----CTGCCCGGCCCTGCCAACAGCAACGCGC 16419
Qy 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
Db 16418 CTCGGGGCCATCAGCTCGAATTCTCGAGCAGTCCAAACGTCGACCTCGCGCCGCAATTC 16359
Qy 471 ValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAla 490
Db 16358 GTGAGAGTATCACCACGACAGAGGTTTCCAGTCCAAACAGCAAGGTCATCAACACCACT 16299
Qy 491 AspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 16298 GACACCATGCTTGAAGTGTCTGTAACATGAGGCT 16263

RESULT 3
AE017180_04
WPCOMMENT
Sequence split into 39 fragments LOCUS AE017180 Accession AE017180

Fragment Name	Begin	End
AE017180_00	1	110000
AE017180_01	100001	210000
AE017180_02	200001	310000
AE017180_03	300001	410000
AE017180_04	400001	510000
AE017180_05	500001	610000
AE017180_06	600001	710000
AE017180_07	700001	810000
AE017180_08	800001	910000
AE017180_09	900001	1010000
AE017180_10	1000001	1110000
AE017180_11	1100001	1210000
AE017180_12	1200001	1310000
AE017180_13	1300001	1410000
AE017180_14	1400001	1510000
AE017180_15	1500001	1610000
AE017180_16	1600001	1710000
AE017180_17	1700001	1810000
AE017180_18	1800001	1910000
AE017180_19	1900001	2010000
AE017180_20	2000001	2110000
AE017180_21	2100001	2210000
AE017180_22	2200001	2310000
AE017180_23	2300001	2410000
AE017180_24	2400001	2510000
AE017180_25	2500001	2610000
AE017180_26	2600001	2710000
AE017180_27	2700001	2810000
AE017180_28	2800001	2910000
AE017180_29	2900001	3010000
AE017180_30	3000001	3110000
AE017180_31	3100001	3210000
AE017180_32	3200001	3310000
AE017180_33	3300001	3410000

AE017180_34 3400001 3510000
AE017180_35 3500001 3610000
AE017180_36 3600001 3710000
AE017180_37 3700001 3810000
AE017180_38 3800001 3814139
Continuation (5 of 39) of AE017180 from base 400001 (AE017180 Geobacter sulfurreducens P
Alignment Scores:
Pred. No.: 7,31e-34 Length: 110000
Score: 693.50 Matches: 175
Percent Similarity: 48.80% Conservative: 70
Best Local Similarity: 34.86% Mismatches: 166
Query Match: 26.82% Indels: 91
DB: 1 Gaps: 13
US-10-009-823A-1 (1-502) x AE017180_04 (1-110000)
Qy 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 49491 GCACGTGTACACCGCATCAGCGGTCTCAACCCCAACGGCGAGGCCATGTCGTCATCGGC 49550
Qy 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGluAsp 43
Db 49551 AACAACTTCCAACTCAACACCATCGCTTTAAGCAGGCGCGGATGCTCTTCTCGGAC 49610
Qy 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db 49611 GTCTCTCCAGCACCATCAGCGGGGGTCC-----CAGATCGGC 49649
Qy 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 49650 CGCGCGTCCAGATCCAGACAGTGGAGAATCAGTTTCAACCCAGGCTCTTTCGAGAGCAC 49709
Qy 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAsp 103
Db 49710 GAGAGCGTACCGACCTCGCCATCCAGGGCGATTCTCTTCGTGGTCCAGAACACACGAG 49769
Qy 104 LysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
Db 49770 GCGCGCTACTATACCGCGCGCGGCTTCTCTTCAATAGGACAAGACCTGTTGAAT 49829
Qy 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
Db 49830 CCGGAGGATATCAGTTCATCGGTACGGCATCTCCCTCGTCGGGACTTTCGACGCGC 49889
Qy 144 ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLys 163
Db 49890 GTGCTCAGCCCATCGATCTG-----ACCACTTTGCCACCATCCGCCGAG 49937
Qy 164 ThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGluSer 183
Db 49938 CAGACTTCCACCGTCAAGTTTCGTGTGAACCTG---GACTCCACCCAGACACCCGCCAC 49994
Qy 184 GluAlaAsnProTyrPheAlaLeuGluSerTrpLysGlyAsnGlyThrProProIle 203
Db 49995 CTGGCG-----TGG-----GACCCCGCAACCCCGTT 50021
Qy 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 50022 GCCACGTCCTTCCCGGATCGCGGAGGAGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 50075
Qy 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 50076 CACACCGCCACGGTGTATTT-----CGCAAGACCCGCCGACACGATGGGAC 50123
Qy 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 50124 TGGCAGGTATCTTCCCGGATCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 50180
Qy 264 GlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThr 283
Db 50181 -----ATCGACGGGACCCCTTCATGTCGACCGGAGCCCTC-----ACC 50222

QY 284 AlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeu 303
Db 50223 GCCAGACTCCCTCGCGCGCGCCAGAACATC----- 50258
QY 304 ValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLeu-----Gln 320
Db 50259 -----ACCTTCGCGGCG-----GGGTCACCGCACCACCCAG 50288
QY 321 ProLeuThrLeuAspPheGlyLeuLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
Db 50289 CCATCTCTCTTCGACCTGGGAGTC----- 50312
QY 341 SerAlaAlaAlaLeuGlyThrAspLeuGlyLysLeuProSerMetMetProLeuGlnThr 360
Db 50312 ----- 50312
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp 380
Db 50313 -----GGCGCTACCAACCCAGTACGCCAGCTCGTGGTGGTTCTTCCAGACCCAGGAC 50366
QY 381 GlyTyrProGlnGlyAspLeuValAspValThrLeuThrSerGluGlyLysLeuGlnGly 400
Db 50367 GGCTACTACAGGGCACCTCCCAAGGTAACCATCGATGACAAGGGATACGTGAACGCG 50426
QY 401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420
Db 50427 GTGTACTCCAAACGCCAGCTTCAGAGCTCTACAGGTGGCGCTGCCAAGTTCCTCC 50486
QY 421 GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440
Db 50487 ACGCCCGGCTGTCCAAAGCGGGTGGCACCCCTCTTCGAGGAGACCTCGAGTCGGGACAG 50546
QY 441 ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
Db 50547 CCCTCTTTCGACGCCAGCACCCCGCGGTGGCGAATCTCGCCAACTCCCTGGAG 50606
QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleLeuGlnArgGlyPhe 480
Db 50607 CAGTCCAAAGCTTGCATCGCGGCCAGTTCGTCAAAATGATCACCACCCAGCGGTGGCTAC 50666
QY 481 GlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500
Db 50667 TCGGCCAACTTCCAAAGAGATCACCACCGCGGACGAGATGCTGCAGGAAGTGTCACTC 50726
QY 501 LysArg 502
Db 50727 AAGCGG 50732
RESULT 4
AF122909
LOCUS
DEFINITION
Treponema denticola Tap1 (tap1), flagellar hook assembly
scaffolding protein FlgD (flgD), flagellar hook protein FlgE
(flgE), flagellar motor protein MotA (motA), flagellar motor
protein MotB (motB), Flil (flil), flagellar switch protein Flim
(flim), flagellar switch protein FlilY (flilY), and flagellar export
protein Flip (Flip) genes, complete cds; and unknown genes.
AF122909
AF122909.1 GI:4426940
ORGANISM
Treponema denticola
Treponema denticola
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
REFERENCE
1 (bases 1 to 9612)
Stamm, L.V. and Bergen, H.L.
Molecular characterization of a flagellar (fla) operon in the oral
spirochete Treponema denticola ATCC 35405
FEMS Microbiol. Lett. 179 (1), 31-36 (1999)
JOURNAL
MEDLINE
95412226
10481082
PUBMED
REFERENCE
2 (bases 1 to 9612)
Stamm, L.V. and Bergen, H.L.
Direct Submission
TITLE

JOURNAL Submitted (22-JAN-1999) Epidemiology, University of North Carolina
at Chapel Hill, CB#7400 2107 McGavran-Greenberg, Chapel Hill, NC
27599, USA
FEATURES
source
Location/Qualifiers
1..9612
/organism="Treponema denticola"
/mol_type="genomic DNA"
/strain="ATCC 35405"
/db_xref="ATCC:35405"
/db_xref="taxon:158"
43..47
-35 signal
-10 signal
64..71
gene
248..1634
/gene="tap1"
248..256
/gene="tap1"
261..1634
/gene="tap1"
/codon_start=1
/transl_table=11
/product="Tap1"
/protein_id="AAD20615.1"
/db_xref="GI:4426941"
/translation="MQALDVRMQALPVKEPERDSGDIKRADAEPRVNGSFLAMIKK
MIAAKDGSKETDEAFEDISLREDKIRDSLOKEAGRNQTELSSEHDHKKKDAENL
LNSKEAYLKEPKNLKKPKTETQKPSKESQAGSKQNDLELNKILNEDFKEIKDF
LPQVDEDEITSLLSDEALXKLDSKSKKQAFDEKTEQAGKGLLSKADKDLKK
ISPOSEAVONLSKKPVSKPKLKISVEDLRSPSVQSDAAIHTTASRSVETONSVM
VIDPSCAQNASQGDQNTQNGSEAKTQSTFSAMLAOIRDASADFVOTKAVLRD
NNGEIKRLHLPNGLKINLELSEKRGVITGVTVASKEAYDAFENLNLEKFPQ
NGEFAELNLSGFGCGEGFAESFSDVFVYKQDRLQLEKTDANLSTSYVYG
PTVDILA"
1632..2155
/gene="flgD"
1632..1640
/gene="flgD"
1646..2155
/gene="flgD"
/codon_start=1
/transl_table=11
/product="flagellar hook assembly scaffolding protein
FlgD"
/protein_id="AAD20616.1"
/db_xref="GI:4426942"
/translation="MQVNNVNNMITTALSOAEMMKDFKSEMPSEKALLGLQVDTLN
KNFAETRVKQKQKDDFLQLLAQLHQDTPSPMEDTQFIGMAQFSSLEQMTNN
KNFAALNMLTGSANVAVGKVDLDLGSSQVSGYISATRGINPEYVWNGVYVWSA
VKTVDNN"
2154..3559
/gene="flgE"
2154..2159
/gene="flgE"
2168..3559
/gene="flgE"
/codon_start=1
/transl_table=11
/product="flagellar hook protein FlgE"
/protein_id="AAD20617.1"
/db_xref="GI:4426943"
/translation="MMRSLFSGVTGMQHMQRMDVIGNVANVNTTGFKEGRVNFQDL
ISQOLSGARPTBELGVNPKVGLGMVASIDITFTQALQTTGVNTDLAIQNGFF
ILKDGKTFVTRAGAFGIDKEGLVNPANGVRVQGMWAEAEGRFINISQSDNLNI
PIGKLDKAKATTSVNYACNLDRKLPELPEGANRAQILSTWSTFEKVVYDSFGEAHLQ
IDFARVGEVNNARATVNVDPNADATATVIGITTDGVONSFIIVRFDNNGHLASVTD
TAGNTPAQVLVQISINVVANGPDEAGAPTRHTFDVNLGEIGTSKNTITQPSDKST
TKAYEQDGYTLGLENFRIDPSGLITGVYNGVRQEIQTAMAGAFANQGLEKAGQNT
YVQSNNSGIANVSTSGTVKGYGTFGGTLEMSNVDLTDFQFVDMIVTQFGQAGAKTIQT
SDTLMETVLNLR"
3573..3575
RBS
3585..3782
CDS
/note="orf4"
/codon_start=1
/transl_table=11

```

/feature="unknown"
/protein_id="AAD20618.1"
/db_xref="GI:4426944"
/translation="MIQVRLNGTKYWINPHQIETIECNPDVTLQMLSGKYVYIKERP
EEILEAVAYRRKIGIFKNEL"
gene
3793..4583
RBS
3793..3796
CDS
3804..4583
/gene="motA"
/codon_start=1
/transl_table=1
/product="flagellar motor protein MotA"
/protein_id="AAD20619.1"
/db_xref="GI:4426945"
/translation="MDIASFIIGFAGIIVAFGAILGGAGGIIHPASPMFIIIGSYM
CLFLTYPLSYTIGIFKVARVFKVTDYGEKALVQRFVSLSEKSRAGLLALEEIEDF
EDPMRGLRNVDGIDGEARFISLMENELNQEERHTWISLVNAWATLAPFGMLGT
VIGLMLLDDKLTALGPNATLVTFYGLMQLNLLVPIATKLMYQNNMEVRSKE
MILEGVIGIAGDNPRILAQRLITYLTPSDRKAIEAEVLK"
4575..5304
RBS
4575..4579
CDS
4585..5304
/gene="motB"
/codon_start=1
/transl_table=1
/product="flagellar motor protein MotB"
/protein_id="AAD20620.1"
/db_xref="GI:4426946"
/translation="MARKKKHGPSAAGSGWLTITADMTIMLCFFVWLPEPSEVDVTO
LMASIGSDPTGGGLSVSAGRLSDLNTISSPMSEKMGKMLASALKXAVLFLAPEI
KTKNIATVDERGIGVLSLSDVFFYPSAELNIAESRDITLLNLQPLSSQDLASRER
VEGHTSDGVDPNVKSNWELSSARAINILHSLTDFGQAESRFSVAGYADTRPISND
TAGRAYNRVDDIILLDDAHF"
5334..5892
RBS
5334..5340
CDS
5347..5892
/gene="flil"
/codon_start=1
/transl_table=1
/product="Flil"
/protein_id="AAD20621.1"
/db_xref="GI:4426947"
/translation="MADNDLMDDDDIQENMASSVDTKRAGLIPMLIKWVAIVLVAL
IFIVTVVIMNIRDKGSHSISPVSEVRETRDVLQYQALGILLKVTADRIPATL
IVDVALGIVNNKSTQESLARKVEIIDFLRSYFKGKTAELRQEKIKIEKHIND
NVLTKNKIDVRFTQDIVEQ"
5892..6941
RBS
5892..5896
CDS
5907..6941
/gene="flim"
/codon_start=1
/transl_table=1
/product="flagellar switch protein Flim"
/protein_id="AAD20622.1"
/db_xref="GI:4426948"
/translation="NTEVLSDDEIDOLLTAISGGDDTDFRAVNDTRKIKIYDKRP
DKESQOMRTVQMHFTFARLTTSLSAQLRSMAHVHVAIVTEQLTYEEFIRISPIPTI
LAINMDPLRASALLIDPSVTISIIDLRFGGQSKVQRELTIESSVMEGVIRTI
LANRRAWTIVDLRFLNSIDTNPQVQIVTFSENVLLVLTETKVQEEGWMNCLP
YITLEPISLISLQFVPSVRRASTQYAAAIKDKLSTVEDVMAEVSGLDVISIDL
NLRAGDVRLPNVRVDPFKLTVGSRPKFCQPGVGRKGLAVQILEKIEDISGDBFE
LTSEGDELVE"
6919..8115
RBS
6919..6924
/gene="fliy"

```

```

/gene="fliy"
6934..8115
RBS
6934..8115
CDS
6934..8115
/transl_table=1
/product="flagellar switch protein FliY"
/protein_id="AAD20623.1"
/db_xref="GI:4426949"
/translation="MSDGSISQDEIDALLSGVGGGIASAPAAAGDDLAGFKKALL
QFTRENPISUNLESSTKGSVSEPIVFSRSEFLRKVSEWTATLIDFSGSMG
DHAFMSPETAKIIVSLVNHEDNVEIDDMALSIVSETIAQYVGTETUSYLERAGITGVA
SSPAESHVPKAMMLPQRFNFYSITNVKLLDSAYQLWEILSEDDVDVKIATAGDPD
SMQGMAGGQNNMGAMQNNMGQMGGATQTGGPAQOMFGSGMPQSGSQGNGQA
MNNMMPMGMGASVOPVPPFQGVSOEBOGNIGLIMDVYMETVELGRTRMKIK
EILGMBGHIIELDKLAGEPVDVLVNHKPIAKGEVVVIBESFGRITTEILSPAERISD
I"
8128..8130
RBS
8141..8800
CDS
8141..8800
/feature="orf10; similar to Treponema pallidum orf3 encoded
by GenBank Accession Number U36839"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAD20624.1"
/db_xref="GI:4426950"
/translation="MSLGGKLFIIIFLFSVTLFAENDGTSDDGLSPSSGILLDAG
KKGNIERNEDGALPAGNSAFDLNVAERTQSPISRLQVLVSLIIVICILAYVVLFLK
KSLFSFSDSPYLKSVASINIAQKSHIVITLGEKAYIVGTVDASINIMIGVEDKTV
DTNNLDAERRSSSPKQDFASMLSSVFKGSKNDVDVNFPAQERLNLKAQAQVPEEK
E"
8787..9612
gene
Alignment Scores:
Pred. No.: 9,51e-29 Length: 9612
Score: 595.50 Matches: 167
Percent Similarity: 46.15% Conservative: 73
Best Local Similarity: 32.12% Mismatches: 205
Query Match: 23.03% Indels: 75
DB: 1 Gaps: 13
US-10-009-823A-1 (1-502) x AF122909 (1-9612)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 2168 ATGATGAGATCTTATTTTCGGGTGTAACCGAGTACAAATACACAAACAAGATCGAC 2227
QY 21 ThrValSerAsnAniIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 2228 GTTATCGGAATAACGTTGCCACGTAATAACACAGGTTTAAAGAGAGGAGAGTAAC 2287
QY 41 PheGlnAspLeuPheSerGlnAspLeuAla-----IleGlySerThrGly 55
Db 2288 TTTCAAGACTTAATTCACAGCAATTAAAGCGGTGCACCCGCTCTCCAGAAAGACTCGC 2347
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValaThrIlePhe 75
Db 2348 GGTGTAAACCCCAAGAGAGTCGGCTTCGGTATGATGTCGCAAGTATTGTACATCTTT 2407
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
Db 2408 ACTCAAGAGGCTTGCAGACCAACACAGGTGTCAACACGACCTTCGAATTCAGGGAACGGC 2467
QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Db 2468 TTTTTCATCTTAAGACGGAGAAAACCTTTTATACAGGCCGCTGCTTCGATC 2527
QY 116 ThrGlnAspGlyPheLeuAsnAspPro---SerGlyPheThrLeuMet----- 130
Db 2528 GATAAAGAGAACCTTGGTGACCCCGCAACGGAATGAGACTCCAGGCTCGATGGCT 2587
QY 131 -----GlySerArgIleSerAsnProAsnIleLysLysGluThrLeuGlu 146
Db 2588 CAAGAAGCTGAGAGATTCAGGATATCAATACATCGGCCCGCACCGGAA----- 2635

```



```
QY 147 ProIleGlnLeuAspPheAsnSerProThrValAlaLysSerProAlaLysThrSerThr 166
Db 2636 -----GATTAAATATTCCTATAGGTCAAAACCTGATGCAAGGCAACCCACA 2693
QY 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
Db 2684 AGTGTAACTATGCTGTGTAACCTGATAGAGGCTGCTGAATGTCGCCGAGGACAAAC 2743
QY 187 ProTyrPheAlaLeuLeuGluSerThrLysGlyAsnGlyThrProPheLysThrSer 206
Db 2744 CGGGCTCAAAATTTGGAATCTACTCTG----- 2770
QY 207 AsnTyrSerThrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
Db 2771 -----TCTACCGCAATTTAAGTATATGACAGCTTCGGTGAACCCACGAGCTG 2818
QY 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
Db 2819 CAAATCGATTTTCCAGAGTTCGGGGGAAGTAATGCT-----TGGCGGCAACCCGTA 2872
QY 247 AlaMetAsnProSerGluAspGlySerAlaAlaSer-----GlyThrAspSer 262
Db 2873 AATGTCGATCCGCAAAATCGAGCGCTACAGTACCCGTGTAGGATTTGGAACA---ACC 2929
QY 263 AlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMet 282
Db 2930 GACGGCGTTCAAAACAGCTTCATAGTCCGTTTGTGACAAATAAGCGGACATCTGCTCCGTT 2989
QY 283 ThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaPro 302
Db 2990 ACC-----GACACTCGAGAAACGTAACG-----TCTCCC 3019
QY 303 LeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeu 322
Db 3020 GCAGGCAAGTTTGTGTTCAAATTTTCATACATGTAGTAGGCGCAACCTGATGAA--- 3076
QY 323 ThrLeuAspPheGlyLeuLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342
Db 3077 -----GCCGGAGCTCCC---ACAAGA 3094
QY 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362
Db 3095 CATACATTTCAGCTAACTTAGCGAA-----ATCGGAATTCCTCAA 3136
QY 363 GlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyr 382
Db 3137 AACACCATTAATCTCTCCGATAAGAGCACTACAAAGCGCTATGCAACAGACGGATAT 3196
QY 383 ProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyr 402
Db 3197 ACTCTAGTTATCTTGAACACTTTAGAAATCGACCAAGCGGTATATTACCGCGGTTTAT 3256
QY 403 SerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAsp 422
Db 3257 TCCAAACGGGTAAAGCGGAGAAATAGGTAGATAGTATGCGCGGTTTTCCTCAATCAGCGC 3316
QY 423 GlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyProGlu 442
Db 3317 GGTCTTGAAAACGGGCAACAAACACTATGTTTCAGTCCAAATACGCGGTATTCGCAAT 3376
QY 443 PheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSer 462
Db 3377 GTTTCACATCGGGAACCGGTAGTAGGAGGCTACTTTATCGCGGGAACACTCGAATGAGT 3436
QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMet 482
Db 3437 AACGTAGATTAAACCGCAACCAATTTGTAGACATGATGCTTACTCAAAAGAGCTTTCAGGCA 3496
QY 483 AsnSerLysSerValThrThrAlaAspThrMetLeuGlnGlyAlaLeuGluLeuLysArg 502
Db 3497 GGTGCTAAACAATACAAACTTCGGATACATGCTTGAACACAGTCTTGAACGAAACGA 3556
```

RESULT 5
AE017255/c
LOCUS
DEFINITION
Treponema denticola ATCC 35405, section 10 of 10 of the complete genome.
ACCESSION
AE017255 AE017226
VERSION
AE017255.1 GI:41819448
KEYWORDS
SOURCE
ORGANISM
Treponema denticola ATCC 35405
Treponema denticola ATCC 35405
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
REFERENCE
1 (bases 1 to 123663)
Seshadri,R., Myers,G.S., Tettelin,H., Eisen,J.A., Heidelberg,J.F., Dodson,R.J., Davidson,T.M., DeBoy,R.T., Fouts,D.E., Haft,D.H., Selengut,J., Ren,Q., Brinkac,L.M., Madupu,R., Kolonay,J.F., Durkin,A.S., Daugherty,S.C., Shetty,J., Shvartsbeyn,A., Gebregeorgis,E., Geer,K., Tsegaye,G., Malek,J.A., Ayodeji,B., Shatsman,S., McLeod,M.P., Majs,D., Howell,J.K., Pal,S., Amin,A., Vashisth,P., McNeill,T.Z., Weinstein,G.M., Norris,S.J., Fraser,C.M. and Paulsen,I.T.
Comparison of the genome of the oral pathogen, Treponema denticola, with other spirochete genomes
Unpublished
2 (bases 1 to 123663)
Seshadri,R., Myers,G.S., Tettelin,H., Eisen,J.A., Heidelberg,J.F., Dodson,R.J., Davidson,T.M., DeBoy,R.T., Fouts,D.E., Haft,D.H., Selengut,J., Ren,Q., Brinkac,L.M., Madupu,R., Kolonay,J.F., Durkin,A.S., Daugherty,S.C., Shetty,J., Shvartsbeyn,A., Gebregeorgis,E., Geer,K., Tsegaye,G., Malek,J.A., Ayodeji,B., Shatsman,S., McLeod,M.P., Majs,D., Howell,J.K., Pal,S., Amin,A., Vashisth,P., McNeill,T.Z., Weinstein,G.M., Norris,S.J., Fraser,C.M. and Paulsen,I.T.
Direct Submission
Submitted (29-JAN-2004) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 123663
/organism="Treponema denticola ATCC 35405"
/mol_type="Genomic DNA"
/strain="ATCC 35405"
/db_xref="ATCC:35405"
/db_xref="taxon:243275"
120..1004
/locus_tag="TDE2679"
120..1004
/locus_tag="TDE2679"
/note="Identified by similarity to ONI:NTL01BH1587"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAS13196.1"
/db_xref="GI:41819449"
/translation="MKKATFICLSVLNLFSCGKKAAETGKSTRAESYWAESKSSSDSAFKEDKESDAVNTASLDNDNIERKLKTGTFETEDIKKTRETIENLVTKYQAVISQEDHQFSQINRQISIRIPKPNFDNLNLTIGIKLDNKNITVEDVTEFVNLRLAKVKEQYTKILSQAKTVKIDLEQVQIDLRSEIEAEGRLYLQKSVYSLTNISMYQIINSSITARPSPFTKALNAVKEGILGFLPSDIIIGILYILWIFILIIAVALIKKRNKIKKSLDN"
complement(1001..1624)
/locus_tag="TDE2680"
120..1624
/locus_tag="TDE2680"
/note="Identified by similarity to SP:083709"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAS13197.1"
/db_xref="GI:41819450"
/translation="MMTVIAVLVLNLFIFFFYKFKNQFSDRASINNKNETOKLVAQAFQDQSDVIMEDKIVNSVLKDRILLACKBEKESAAERVYESLAATAKANAQFGDPDAFVKTKPVISNESVKETRSIEVNKENPKNIDNEPKIYTKQLAPKSGIVPDQATASKQIEMAKKGLSIELIAEKVPLPTGCEIILISMNT"
complement(1621..1959)

```

/locus_tag="TDE2681"
/complement(1621..1959)
/locus_tag="TDE2681"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS13198.1"
/db_xref="GI:41819451"
/transl_table="MAIQPDLQTLTYAQMAKVQKQGAEOQLASWVKDAQOQONKIDA
EKLSTVLQVDSNNLNRKGRSPDENKGGHPKQNTAQIQEPPDSSEYIKDP
SLGRVDISG"
/complement(1940..3910)
/locus_tag="TDE2682"
/complement(1940..3910)
/locus_tag="TDE2682"
/notes="identified by similarity to OMNI.TP0710; match to
protein family HMM PF03961"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAS13199.1"
/db_xref="GI:41819452"
/transl_table="MVLNQIQKMGEMHELDGRFFVFDISGTLDEALANAAIQGM
PVSIDYIILQKAGSGFFAIVPKWEKIRAYETVKKPQNIIDEKVETEDIEDGTVI
KDGMYVYCAADGIVPKVAPGTGHAPDIKAARKFRDRALPIPEDDILAPILPEEK
TGEVYVAPYKRIIPGNDAMVNIISDEMKAIVLYVPTPTGGVLSADTIIAFLKNR
IIVGNEERVQKQDSVPYREDLYAEGIAPQNGADAKIILYNEPVDNTRVLQETESG
QINFKEINLQNVVEGQVQAKVPAQRGKAGTIVTKYLEALNGKDVAMPVGNKTKIA
ADGLTIVAGNGQVLVKNKIVTOEIVYVEGDVSIKNTIFLGSVFNVDGDFVI
KASNIIVKSGRAEDTEDIIVSQGINGKEGVIRACKSIWSKFIQNTDVRAGD
MVIYDSGIIKSNVWANKIICRGKADIIISGNLSAGESISARNLGSVSGNDLMSVG
FDPKSRERLFLQKQMDQKSLDNLNLSLEKLSKRGELPKDKEENYNKMEYK
YTLQTDHEVQKQITQIKYLNLTNKGVSASGHVYVGVRIIVIRDTEDVRMDCKAT
TFYLDKIVRYGYQENEDPKKVPYSGYSTN"
/complement(3919..4707)
/locus_tag="TDE2683"
/complement(3919..4707)
/locus_tag="TDE2683"
/notes="identified by similarity to SP:P31804; match to
protein family HMM PF04539; match to protein family HMM
PF04542; match to protein family HMM PF04545"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-27 factor"
/protein_id="AAS13200.1"
/db_xref="GI:41819453"
/transl_table="MANTDENIPBELMDKYYKTSDPKIREYFIKYAPLVKYVAGK
VIGMPNPFDDLVGVGGLLDIALEKDDLNKRVNFYAVNRIGALPDEIRSIDW
VPSVRQKSEIETADLEARLGRSASNEIAYAGNDIEYNSLLKISATSVISL
TDLRFSDSDSEFVSVDIIEAPSLNPDVIVEREDIKRIVIEAKELPDEPKVLTIMY
VYEDMTREIGEVHLTVESVSQIHTSANIKLRAKLSNVTGKIK"
/complement(4732..5400)
/locus_tag="TDE2684"
/complement(4732..5400)
/locus_tag="TDE2684"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS13201.1"
/db_xref="GI:41819454"
/transl_table="MAKFRPLICTTAALALSFLGLISGVRFTLLRLSLISLISG
MIFALPQDLDHFDLPQVNAEKAETGKNLISIDEPVMPFIDDEYDPESD
MIEAPYVDESISDDATLDEIDENIKDDSQASKNVSGSGTEKQKQNTV
KELELPDLQDFTPEDPEQDQMDFTQGTSAFVSTDLASGMDNTWNAIRTVL
KRDS"
/complement(5448..6335)
/locus_tag="TDE2685"
/complement(5448..6335)

```

```

/locus_tag="TDE2685"
/notes="identified by similarity to OMNI.NTL03PA01455;
match to protein family HMM PF00991"
/codon_start=1
/transl_table=11
/product="flagellar synthesis regulator FleN"
/protein_id="AAS13202.1"
/db_xref="GI:41819455"
/transl_table="MTQADSLTKLMKNKSSKAVLSDLPPPKRSRIIAVTSKGV
KTNISNMAIYAKMGKIVIVDADIGLANVNMNIIPFNLYHVMKQKQKSDII
IDTYGKFKVAGSGFKIANMEAEESDFIKELYTLAEADIIIDTSAGVKNVLSF
VAADEVIVITTSPTAITDAYGIKIIATEVENYDLNLKMWVNRVNSALEGKIIAR
MIQVAFQNLKQVYLGFIYNDPAVEQAVLKQKPFISAPKSKAASCLRHIVAKLEKT
DYNEYSGLSGFLQKLFGRK"
/complement(6337..7677)
/locus_tag="TDE2686"
/complement(6337..7677)
/locus_tag="TDE2686"
/notes="identified by similarity to SP:Q56339; match to
protein family HMM PF00448"
/codon_start=1
/transl_table=11
/product="flagellar biosynthetic protein FlhF"
/protein_id="AAS13203.1"
/db_xref="GI:41819456"
/transl_table="METFVEASTYKCIQIOEKYGNIMITRREKRNGLSLFN
KETRLSFNQNEAFVFKSNNEEVEKSAVAPIKEYSPANLAELKTMKLAASQS
EEMAKMPPYKLENGESQPIKIDIESKFKLAETVERLVEIRLKNPSQHEHNI
SKAALLENDFETAKYIRSIKESIKSNLAEIENDFELVOKKVLWDIASSIQKLEET
NNDPSKKQKIALVPGTGTITLAKLAAYILAVSKERSLDVRVITIDYRIG
AAFOIKKYCHMGPIIATDPLDKYLDKSDADIICTIDTSPTPEKILENQ
KYPEIPBGRITETHLVVSATKAADIIEIKQSVFNFSLLITKIDETAHVGSIIIV
LDEYKIPVATITGQTVPKDIARASKVLRLKLTGFLDYNENFNDSEISVNS"
/complement(7748..8806)
/locus_tag="TDE2687"
/complement(7748..8806)
/locus_tag="TDE2687"
/notes="identified by similarity to GP:1895094; match to
protein family HMM PF00155"
/codon_start=1
/transl_table=11
/product="L-threonine-O-3-phosphate decarboxylase,
putative"
/protein_id="AAS13204.1"
/db_xref="GI:41819457"
/transl_table="MEHGVLYKDKDEVLIIDFSSNINPLMAPKGLKALSASFNLL
VYPIRVSLLKVTAYOYLKCRPENIVLGNAGVEIIDNFCSMKRVVLCTPCFSEYGLR
ALVHNKPVLELPYLPDPLDGLAGLEKLRAGDILLGNPNPTGLRLEKTLALQYSL
VQTKFVLLDDEAEYECFPDYDSIALPKDGYKNICIIAAKFKFGLPGLRIGYACT
STDFAALSKEMAWHINAFEAAPAFDFEDFIKASKAYIOKRVFLLENIEKYGT
SGEIKLQTKYQCNFILLKLNKADKEDALFEKKGILIRTCSSFKTLGDNHRIAVR
SHKDNCKFINAISRSKSK"
/8859..8984

```

Alignment Scores:

Pred. No.:	1.96e-27	Length:	123663
Score:	595.50	Matches:	167
Percent Similarity:	46.15%	Conservative:	73
Best Local Similarity:	32.12%	Mismatches:	205
Query Match:	23.03%	Indels:	75
DB:	1	Gaps:	13

US-10-009-823A-1 (1-502) x AB017255 (1-123663)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db	103062	ATGATGAGATCATTTATTCGGGTGTAACCGGAATGCAAAATCAACAAACGAATGAC 103003
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40

```
Db 103002 GTTATCGGAATAACGTTGCCACGCTAAATACACAGGTTTTTAAGAGAGGAAGTAAC 102943
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAla-----lleglySerThrGly 55
Db 102942 TTTCAAGACTTAATTTTCAACAGCAATTAAGCGGTGACCGCGTCTACCGCAAGAACTCGGC 102883
Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValargThrIlePhe 75
Db 102882 GGTGTAAACCCCAAGAGTCGGCTTTGGGTATGATGTCGCAAGTATTGATCAATCTTT 102823
Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
Db 102822 ACTCAAGGAGCTTGCAGACCAACAGGTGTCAACACCGACCTTGCAATTCAGGGAACGCG 102763
Qy 96 PhePheGlnValThrLeuGluAspLysValHisThrArgAlaGlyAsnPheargPhe 115
Db 102762 TTTTTCATCTTAAAGACGAGAGAAAAAATTTTATACAGGCGCGTCTTCGGTATC 102703
Qy 116 ThrGlnAspGlyPheLeuAsnAspPro---SerGlyPheThrLeuMet-----130
Db 102702 GATAAAGAGGACCTTGGTGAACCCCGCAACGGAATGAGAGTCCAGGCTGGATGGCT 102643
Qy 131 -----GlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu 146
Db 102642 GAAGAAGCTGAAGGATTCAGGATAATCAATACATCGGGCCAGACGGAA-----102595
Qy 147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
Db 102594 -----GATTAAATATCTCTATAGGTCAAAAATCTGATGCAAGGCAACCA 102547
Qy 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
Db 102546 AGTGTAAACTATCTGTAACTTTGATAAGAGGCTGCTGAAATGCGCGAAGGAGCAAC 102487
Qy 187 ProTyrrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProIleSerThrSer 206
Db 102486 CGGGCTCAATTTTGGAACTACTTGG-----102460
Qy 207 AsnTyrrSerTrpAlaGlnProMetArgValTyrrAspGlnGlnGlyAsnSerHisAspIle 226
Db 102459 -----TCTACCGCAATTTAAGGTATATGACAGCTTCGGTGAAGCCACCGAGCTG 102412
Qy 227 ThrValTyrrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrrLeuVal 246
Db 102411 CAATCGATTTTCCAGAGTTCCTCCGGGAGTAAATAGCT-----TGGCGGCAACCGTA 102358
Qy 247 AlaMetAsnProSerGluAspGlySerAlaAlaSer-----GlyThrAspSer 262
Db 102357 AATGTGATCCGACAAATCAGACGCTACAGTACCCGTGTAGGAATTTGGAACA---ACC 102301
Qy 263 AlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMet 282
Db 102300 GACGGCGTTCAAAACAGCTTCATAGTCCGTTTTCACATAAGCGACATCTTCTCTCCGT 102241
Qy 283 ThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaPro 302
Db 102240 ACC-----GACACTGAGAAACGTAACG-----TCTCCC 102211
Qy 303 LeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeu 322
Db 102210 GCAGGGCAAGTTTGGTTCAATTTTCAATATGATAGTAGGCGCAACCTGATGAA---102154
Qy 323 ThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342
Db 102153 -----GCCGAGCTCCC---ACAAGA 102136
Qy 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362
Db 102135 CATACATTTGACGTAACCTTAGCGGAA-----ATCGGAACCTTCCAAA 102094
Qy 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrrSerGlnAspGlyTyrr 382
Db 102093 AACACCATTAAGTCTTCCGATAAGACGACTACAAAGCGCTATGAACAGGACGATAT 102034
```

gene
/gene="murF"
/locus_tag="Bd3204"
complement(675..2096)
CDS
/gene="murF"
/locus_tag="Bd3204"
/EC number="6.3.2.10"
/function="UDP-N-acetylmuramyl pentapeptide synthase"
/note="InterPro: Cytoplasmic peptidoglycan synthetases
N-terminal
hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="UDP-N-acetylmuramoylalanyl-D-glutamyl-2,
6-diaminopimelate-D-alanyl-D-alanine ligase"
/protein_id="CAE78024.1"
/db_xref="GI:39576795"
/translation="MRAMDQTIIVKVTGAKILSKQETNFAIGTDTRAQLOQLFIAL
KGAFDAHFLDKAVEQASGLLVHEENALIEKDKATILLYPDLTKALQKLNWAR
HESARIVGITGSGKTTTKEFTAAALVGSMDVHNGFNNHGWPFLLQLPNNKD
VAITEMGNHAGETTELVAHAEEDVVVCTWVGRHMEFFETIEKVAEAEKEETIEAAGD
QTVRIYLDNTQTHMYKADKEPKARILTESSEPRADVHLMISSMMNSISLAKGS
IGGARFARVQVGAQNLNMAASVGLAVGNTPAQVWAGLPAKTNNGRQLVHLK
SGAQMFDAIPDPSNMKLLIDNNKLLTVPRKRVGVFGQNRGSGASANLHEELGWV
GQAGFDKVFYIGDDADAFAGLSSGAKYKALIEKDYKSSGODLAGFLTKGTIAVVK
ASRGTKLERFVFCEPLDFAEQ"
complement(2099..3589)
/gene="murE"
/locus_tag="Bd3205"
complement(2099..3589)
CDS
/gene="murE"
/locus_tag="Bd3205"
/EC number="6.3.2.13"
/function="UDP-N-acetylmuramyl tripeptide synthase"
/note="InterPro: Cytoplasmic peptidoglycan synthetases
N-terminal
hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="UDP-N-acetylmuramylalanyl-D-glutamate--2,
6-diaminopimelate ligase"
/protein_id="CAE78025.1"
/db_xref="GI:39576796"
/translation="MKLQHLFSILPGIPENAFSHIEVTGVNDARLVVFGSVFAIRG
NKLDHTFIIDAVAKGAALVVEDKAKIEGEGVGVVQVPSREVLVDLASRYWDPG
QELFCVGTGTNGKTSVTYMTAELAHGKIPTGIVGVNHLGDQVWPSEMTTDPFV
LOKRLRFRSAGALAVAMEVSSHALDQRRVDSVPFNTVPTNLTRDHLHYHNTMSYL
EAKORLPDLDLWTKRPPCFATVNTADKFGRRLRVADPAVLWYGEKSDIRYEILKM
DFALTPEKVTVPYGEVEVRLPMSGTHNVNMAALGAGLSAGLPLNLCIEALDRFTGV
PGRLOSPVNDKLSVFVDYHTPDALENVLMATKYRENLOSQSRINTIFGCGGBDK
GRPLMAEMALKYSDVEVITSDNPREDPQTIIDILAGVGGANKAKATIVVDRKDAI
TOTLKAQEGDVLIIAGKHEDYQITGTQKFPFSDVKVAEEALQGR"
3617..4297
/gene="estA"
/locus_tag="Bd3206"
/EC number="3.6.1.1"
/function="predicted esterase"
/note="InterPro: Phospholipase/Carboxylesterase
hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="serine esterase, putative"
/protein_id="CAE78026.1"
/db_xref="GI:39576797"
/translation="MRQGLKHLQEQEINHNDAPWVIFPHGYGADANDFLSLGELIPTK
KTYNWLFPNGLEVPICPAWTKGAWWITDMWEIOAQRGEHRDFNSDTPKMSKAYD
LAMEIIPQMKVPWNKIYVGGFSQACMLATEIYLRAPETPKGLVIMSGTLVPHODEWKQY
VPPNAGORFFVQSHINDAVLGYKQAQKLETLITLQNGMKSLQFGRGHIEIPWVITQI
GEYLNITIP"

gene
/locus_tag="Bd3207"
CDS
/locus_tag="Bd3207"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="CAE78027.1"
/db_xref="GI:39576798"
/translation="MKYLIILAFLLTSFAFAHRLVLEDPSEIQTLSEADPNLGRVLV
SLPGADNKKLLEHPAYNSLVADLEKLEALKTQSKUGVGNITVHRLFLRWLKSPPAA
RFEIVGVNRMDFVAFNPGTCGEVRLIYRLAYTKQNTOTLIQRLPMTLNAAFKLPAT
PNCQAVAKQSSQKQPLKPWVTLANLKSLEVLNLAQVRFSTIGDMGHAETFLRY
RLONNVFASAMETPDVARLQNSALRAELLAWISTENIKALDAGILNIPKFLTY
KSSSFALHGMNRLANRPDQIFKKEDEFNKLTDLENIVGSPFRRLNLDLSGAGCHQ
GRTIAGHFHGLKDPESITIFANSVFSSQSPHMOEQEKEROVYFQNIPOGKAADSRPFS
ERDPTPGAMNACHGLPFGSEYQAWTCAAGLCKMAVYTPEGSKIGECLPFPQWAGNPC
EPGVSTQTNASHGDKMAGATRTSCPSHQYQCTTKVFPFGMCGCGCQLTSGEACGAI
AVLQGFNDCLARHPPFGECLAQNTRPAGLQACDDTISCADYITCNKTQSGKVCIPPY
FLFQLRVDGHPKPI"
4339..4362
/locus_tag="sp0828"
4339..4362
/locus_tag="sp0828"
6108..6650
/locus_tag="Bd3208"
6108..6650
/locus_tag="Bd3208"
/locus_tag="Bd3208"
/function="cAMP-binding proteins - catabolite gene
activator and regulatory subunit of cAMP-dependent protein
kinases"
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="probable cyclic nucleotide binding protein"
/protein_id="CAE78028.1"
/db_xref="GI:39576799"
/translation="MRPHCLHPHAPVTPTRILARWQKSLIITYDPMFNKASKGKXP
IDQIVLKLADGEILFREGDLSREMYIVQKGAVEFRKVEGQTMILGKVDGSMVGENS
LLESLPRASAQAVGETTLLMDFRGSLLKIRRDPTFAFELMKQLSGRIESTNEKLIT
LMAAQLSRNDLQIAESTL"
6647..7402
/locus_tag="Bd3209"
6647..7402
/locus_tag="Bd3209"
/function="Uncharacterized protein conserved in bacteria"
/note="Ser/Thr protein phosphatase family protein
hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="putative phosphatase protein"
/protein_id="CAE78029.1"
/db_xref="GI:39576800"
/translation="MIIDVTTEKMWISDLHLGNPFSEAKSLIEFLYAAENKYDYC
INGGLEIAQVSPRKAEDVPEVPAVKAITKAGRVFYVVGNDHIVLENFLEDWGPL
TLAPLNVRCGKORIRIEHGHLYDPFFVKRDLVEFLTLGGFALWVSPRIYKAWIKP
BELSKRRRRPRGSGSEVPCLPGEHPDVIABEELCSRGFDALFGHTHTFGTIDLPGQ
RIFNSGSMWISPTIIGTISQDVLSQFPAKGA"
complement(7409..9001)
/gene="recQ"
/locus_tag="Bd3210"
complement(7409..9001)
/gene="recQ"
/locus_tag="Bd3210"
/EC number="3.6.1.1"
/function="superfamily II DNA helicase"
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=1

/product="ATP-dependent DNA helicase RecQ"	
/protein_id="CAE78030.1"	
/db_xref="GI:39576801"	
/translation="MKVSFMEIAVSPQENLKKYFNLSFRGQKEIIESVMGQDVL	
AVLPFGGKSLQYPAVATKLVIVISPLIALMKQDVASLRVIGIPAGALHSGSD	
DKREYFAINKGFAVLVLSPEKAQEGFHWQNRVGLFAIDEAHCSQWHDRE	
EYALNVLKLCPLDVPDLTASATPVLDDISKHLKQKPMVHGVRNLYQVE	
LCEDEDAKLILLQLSIKQTPGRIVVCGTRKVTESIAAFLLQKFKGVSQVYHAGLTS	
Alignment Scores:	
Pred. No.:	7,378-26
Score:	Length: 349965
Percent Similarity:	579.50
Best Local Similarity:	44.55%
Query Match:	30.59%
Query Match:	22.41%
Query Match:	119
Query Match:	16
Query Match:	1
US-10-009-823a-1 (1-502) x BX842655 (1-349965)	
QY	1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db	181703 ATTCTTCTTCATGTACACTGGTGATCCGGTATGACAGCAGCGGCGAGCTCGGT 181644
QY	21 ThrValSerAsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValVal 40
Db	181643 GTATCCGGGCAATATCCCAATGGAACATATCGGTTCGAAAGCAAGCCGTCACAG 181584
QY	41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db	181583 TTCCAGGACATCATCTCCAAAATCTA-----AAAGGTATTGTGGTGGTAAAC 181536
QY	61 GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPhe 80
Db	181535 CAAATGGCGCGGTGTGAAGATCGGTGCGGTAATCCCAATCTTCCAGGGTAAACATC 181476
QY	81 GluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThr 100
Db	181475 GACGCCACTGAAAGTGAAGACCTTGCAATCTCCGGTGACGGTACTTCAAGGTAAA 181416
QY	101 LeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPhe 120
Db	181415 GGTTCGCGAGGTGAGTCTACACTCGTACGGTTCCTTCCACTTGCACCGTGAAGGCTAC 181356
QY	121 Leu-----AsnAspProSerGlyPheThrLeuMetGlySerArgTyrSerAsnAsnPro 138
Db	181355 CTGGTAACAAACGAC-----AATCAACGGGTACAGGGTCTCTCT 181317
QY	139 AsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAla 158
Db	181316 ACAGATGAGAGGCAATATCGTCAACAAATGACCGATATCAAAATTCCTCCG---GCT 181260
QY	159 LysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThr 178
Db	181259 CTGATCCCGCTAAAGCAACCAAGAAATTTGAAACTGGATCTGAACCTGGATCCCGCATG 181200
QY	179 AspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSerTrpLysGlyAsn 198
Db	181199 GACCCACCAAG-----AAGTTCGAC 181179
QY	199 GlyThrProProLysThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAsp 218
Db	181178 CCAGCGGATCCTTACTCCACTCTCACTATCCACTGGT-----GTCAGATGATGAC 181125
QY	219 GlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerThrGly 238
Db	181124 TCTCAGGTTAAACCAACCTTGCTGACGCTCTCTCTTC-----AACAAAGTTAAC 181077
QY	239 SerLysThrPheGluTyr-----LeuValAlaMetAsnProSerGluAspGlySerAla 256
Db	181076 GACCGCAATGGAATTCAAAGGTTGGTT-----GACCGTAAAGAG 181035
QY	257 AlaSerGlyThrAspSerAlaGlyLeu-----LeuMetSerGlyThrMetThrPheSer 274

Db	181034 ATCACTGGCGGGAAGGCAAAATGTCTGAAGTGGCTGCTGTAACATGATGTTCAACC 180975
QY	275 SerAsnGlyGluLeuLysAsnMet-----ThrAlaPheThrProThrGlySer 290
Db	180974 GTTGACGGTAAATGGATTCCTCAGGAAACCAACCTCCAGAACTTCAACTTCAAGGTGGT 180915
QY	291 AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPhe 310
Db	180914 GCTTTGAGGAC----- 180903
QY	311 SerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly-----Ile 328
Db	180902 -----CAGCAAGTGAAGTGAACCTCGGTGATCGGATC 180870
QY	329 LysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp 348
Db	180869 AAG-----GAC 180864
QY	349 IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg 368
Db	180863 GCGCGTAAA----- 180855
QY	369 AsnGlySerSerThrArgArgTyrSer-----Gln 379
Db	180854 --GGTTTGACGGTACTAAGCAGTACGGTAAACCTCGGACCTGATCTCTTGGCACAC 180798
QY	380 AspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGln 399
Db	180797 GATGGCGCGCTCGGGTACATCAGCTGCTTGTCTTCAATGACGAGGATCTTGCACA 180738
QY	400 GlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThr 419
Db	180737 GCGGTATCTCCACGGTCAAGCCCAATGACCTTGGCCAGATCGCTTGGCAAGTTTGA 180678
QY	420 SerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGly 439
Db	180677 AACCCAGAGGCTCTGTTCAAAGTTGGTAAACACCGTTTGAAGAAATCCAGAGATCCCGGT 180618
QY	440 GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db	180617 ACTGCTCCGAGTAGTCTCAGGCGCTGCGCGCGGTAACACTGTTGCGCAAAATCTTTG 180558
QY	460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGly 479
Db	180557 GAAAGATCCACTGCTGATCTGGCTACAGAGTTCGTGAACATGATCCAGAAATCAGCGTGGT 180498
QY	480 PheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGlu 499
Db	180497 TTCAGGCGCAATGCCAAACCATCACCAGCAGCGCAACTTCTGAACGAGTATCCAG 180438
QY	500 LeuLysArg 502
Db	180437 TTGAACGA 180429
RESULT 7	
TPU04619	
LOCUS	
DEFINITION	
Treponema phagedenis Kazan 5 flagellar hook polypeptide (figE)	
gene, complete cds, MotA gene, partial cds.	
ACCESSION	
U04619	
VERSION	
U04619.1	
KEYWORDS	
SOURCE	
ORGANISM	
Treponema phagedenis	
Treponema phagedenis	
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.	
REFERENCE	
1 (bases 1 to 2516)	
AUTHORS	
Limberger, R.J., Sliwinski, L.L. and Samsonoff, W.A.	
TITLE	
Genetic and biochemical analysis of the flagellar hook of Treponema	
phagedenis	
J. Bacteriol. 176 (12), 3631-3637 (1994)	
MEDLINE	
94266716	
PUBMED	
8206841	
REFERENCE	
2 (bases 1 to 2516)	

AUTHORS Limberger, R.J.
 TITLE Direct Submission
 JOURNAL Submitted (20-DEC-1993) Ronald J. Limberger, Microbiology,
 Wadsworth Center for Labs and Research, David Axelrod Institute,
 New York State Dept. of Health, 120 New Scotland Ave, Albany, NY
 12201, USA

FEATURES
 source
 1. .2516
 /organism="Treponema phagedenis"
 /mol_type="genomic DNA"
 /strain="Kazan 5"
 /db_xref="taxon:162"
 /clone="phase 3"
 /clone_lib="gt 11 random partial digest"
 <1. .294
 /note="ORF1; homologous to ORF8 of Bacillus subtilis FlaA
 locus"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAA73466.1"
 /db_xref="GI:439980"
 /translations="PAPMEDTQFVQMAOFTSLEOMTVNSQSEKLNLLSGSEAVNA
 VGGVNIEDGSIKVSQVITATRGRIPOVQVDGKWDVSTVTKTVIGEAQNR"
 319..324
 /note="flag site"
 334..1725
 /gene="flgB"
 334..1725
 /gene="flgB"
 /codon_start=1
 /transl_table=11
 /product="flagellar hook polypeptide"
 /protein_id="AAA73467.1"
 /db_xref="GI:439981"
 /translations="MNRSLFSGVSGMQRHQRMDVIGNVVNVTTGFKRGRVNFQDL
 ISQLSGASPNVEGVNPKVGLGVVVASIDTVHOGALOTGTINTDIAIQNGGFF
 ILKDGKSFVTTAGAGVDRDGLVNPANGMVQGMWAEIDISQOIINTSDOTELII
 PIQKIDAKATTIVACHLKRLPELPEGANQADILRSTWADENVDTFGEQHLQ
 MVSFVPTNNQMLATVNDVPEQAGTEIRVIGITDGTENTFIVSFNDYHLSAVTD
 TAGNVAPAGVLQVSYNVVANGPDEGAPTRHTFNINLGEISRTNITQFABRST
 TRAYQDGVAMGYLENFKIDQSGIITGVYNSGASREIGQLAGFANQGLEKAGENT
 YIQSNNGIANITVSGVWGKGLIAGTLEMSNVLDLTQFTDMLITQRFQAGAKTIQT
 SDMLTETVLNLR"
 1740..1783
 /note="Mota gene site"
 2023..2027
 /note="Mota gene site"
 2038..2514
 /codon_start=1
 /transl_table=11
 /product="Mota"
 /protein_id="AAA73468.1"
 /db_xref="GI:439982"
 /translations="MDLASFIFGFGAFAILMGILGGSAGSFHLPVSFIVTGGSYL
 TFLAYPLSYTGIFVKVCARVFSADPHEKEIVQRIYALAEKSRRTGLLAEEIQDF
 DDEFMTGLRNVVDGIDGAIKRLMENELSHMEERHNRWISFINAWATLAPGYMLG"

stem_loop
 RBS
 CDS

Alignment Scores:
 Pred. No.: 1.28e-27 Length: 2516
 Score: 567.50 Matches: 160
 Percent Similarity: 43.95% Conservative: 69
 Best Local Similarity: 30.71% Mismatches: 215
 Query Match: 21.95% Indels: 77
 DB: 1 Gaps: 10

ORIGIN
 US-10-009-823a-1 (1-502) x TP004619 (1-2516)

Qy 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 334 ATGATGAGATCAATTAATTTCCGCGGTTCGGAATGCAAAATCACCAGACAAGATGAT 393

21 ThrValSerAsnAsnAlaAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 394 GTTATCGGAATAGCTGCAACCGTGAATACACAGGTTTAAAGCGCGGAGAGTTAAT 453
 Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 Db 454 TTCCAGGATTTAATTTCTCAACAATTCAGCGGTGATCCCGTCCGAATGAAGAAGTCGCG 513
 Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 Db 514 GGTGTTAACCCGAAAGAGTAGGCTCGTGTATTGTCGCAACGATTGATACGTTACAC 573
 Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
 Db 574 ACACAAGAGCGCTTGCAAACTACCGGAATCAATACTGATATTCGATACCAAGGAACGGA 633
 Qy 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 Db 634 TCTTTTATTAAAGACGAGAGAAAGAGTTTACACACAGCAGGTGATTCGGTGT 693
 Qy 116 ThrGlnAspGlyPheLeuAsnAspProSer-----Gly 126
 Db 694 GACAGAGACGCGACATTTGGTAATCTCTGCCAACGGAATGCGCGTGCAAGTTGGATGGCG 753
 Qy 127 PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlyThrLeuGlu 146
 Db 754 GAAGACATTTAGGGGCGACCAATTAATCAATACATCCGATCAACCGAAGACTTAATTATT 813
 Qy 147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
 Db 814 CCGATTGCT-----CAAAAATTGATGCAAGGTGTACCAACC 849
 Qy 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186
 Db 850 GATGTTGCTGATGCGTGCAACCTTGATAAAAGCTTTACCCGAATTCGCCGAAGGTGCAAT 909
 Qy 187 ProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProIleSerThrSer 206
 Db 910 CAACGGATATTTACATCACTGCG-----936
 Qy 207 AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
 Db 937 -----GCAACCGATTTTAACTATACGATACCTTTCGGGGACCAACATAAATCG 984
 Qy 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
 Db 985 CAGATGGTTTCTCAAGGTCCTCCGCTGACAAACAAC-----CAGTGGCTCGCA 1032
 Qy 247 AlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeu 266
 Db 1033 ACCGTAATGTTGATCCGAAATCAGCGCGGAACCGAAACCGGTGTAGGTATTGGAACA 1092
 Qy 267 MetSerGlyThr-----MetThrPheSerSerAsnGlyGluLeuLysAsn 281
 Db 1093 ACCGACCGAAGCAAAATACCTTTATTGTCAGCTTTGACAACTATGGGCATTTGGGCTCG 1152
 Qy 282 MetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAla 301
 Db 1153 GTTACC-----GACACGCGCGGAATGTTACA-----GCG 1182
 Qy 302 ProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnPro 321
 Db 1183 CCGCGCGGACAGGTTTATGTCAGGCTTCTTATAATGTTGTCGCGCA-----1230
 Qy 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAlaSer 341
 Db 1231 -----AACCTGATCAAGCGGTGCTCCACT 1257
 Qy 342 AlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSer 361
 Db 1258 CGTCATACCTTTAACTCACTCTTGAGAA-----ATAGAACATCT 1299
 Qy 362 SerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGly 381

Db 1300 CGCAATACGATTACACAGTTTCCGAAAGAGCACTACAAAGCATATCAGCAGACGGT 1359
QY 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 1360 TAGCAGATGGCTATCTTGGAATTCAAATTTGACCAAGTGGTATTATCACCAGCGTA 1419
QY 402 TyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGlu 421
Db 1420 TATTCAACCGGTCAAGCCGTAATCGGGCAGCTTGCTTGGCAGGTTTGGCAACACAG 1479
QY 422 AspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyPro 441
Db 1480 GCGGTTTGGAAAGAGCAGGGGAAACACTTATTCAATCGAACACTCGGGTATTGGC 1539
QY 442 GluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThr 461
Db 1540 AACATTACCGTATCAGAGTTATGGTAAAGAAAGCTTATTCCAGGAAACACTTGAATG 1599
QY 462 SerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGln 481
Db 1600 AGTAACGTGGAATTAACCGATCAATTTACCGACATGATTATTACAAAGAGGGTTTCAG 1659
QY 482 MetAsnSerIlySerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 1660 GCGGAGCAAAACCACTTCAAACTTCGGATACATCTCGAAACTGTGTGAATTGAAA 1719
QY 502 Arg 502
Db 1720 CGG 1722

RESULT 8

CR522870.30.

WPCOMMENT

Sequence split into 36 fragments LOCUS CR522870 Accession CR522870

Fragment Name	Begin	End
CR522870_00	1	110000
CR522870_01	100001	210000
CR522870_02	200001	310000
CR522870_03	300001	410000
CR522870_04	400001	510000
CR522870_05	500001	610000
CR522870_06	600001	710000
CR522870_07	700001	810000
CR522870_08	800001	910000
CR522870_09	900001	1010000
CR522870_10	1000001	1110000
CR522870_11	1100001	1210000
CR522870_12	1200001	1310000
CR522870_13	1300001	1410000
CR522870_14	1400001	1510000
CR522870_15	1500001	1610000
CR522870_16	1600001	1710000
CR522870_17	1700001	1810000
CR522870_18	1800001	1910000
CR522870_19	1900001	2010000
CR522870_20	2000001	2110000
CR522870_21	2100001	2210000
CR522870_22	2200001	2310000
CR522870_23	2300001	2410000
CR522870_24	2400001	2510000
CR522870_25	2500001	2610000
CR522870_26	2600001	2710000
CR522870_27	2700001	2810000
CR522870_28	2800001	2910000
CR522870_29	2900001	3010000
CR522870_30	3000001	3110000
CR522870_31	3100001	3210000
CR522870_32	3200001	3310000
CR522870_33	3300001	3410000
CR522870_34	3400001	3510000
CR522870_35	3500001	3523383

Continuation (31 of 36) of CR522870 from base 3000001 (CR522870 Desulfotalea psychrophil

Alignment Scores:

Score: 2.21e-25 Length: 110000
Matches: 563.00
Percent Similarity: 39.87% Conservative: 75
Best Local Similarity: 27.81% Mismatches: 190
Query Match: 21.77% Indels: 184
DB: 1 Gaps: 18
US-10-009-823A-1 (1-502) x CR522870_30 (1-110000)
QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 27852 GCATTATACACAGGAGTTCAGGGACCTTAACACAGCCAGCCAGCAATGATGATCAAG 27911
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAsp 43
Db 27912 AACAACTTCGCAACACACACACTATTGCTTATAGGATCTCGTACAAATGTTTCAGAT 27971
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlnGlnValValPheGlnAsp 63
Db 27972 CTCATTATCTCCAGCGTC-----ACTGGTTCGGGTGGCAGCTCTCAAGTAGGC 28019
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 28020 CGTGGGTTGGCTCTCCGAGTGGACACAGATTTCAGTCAGGTACATTTGATCAACC 28079
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGlnVal-----ThrLeu 101
Db 28080 TCCTCCGAGACTTGATCTGGCCATTGAAGGAAACAGCTTTTCTCTCAGCCCAAGCAAC 28139
QY 102 GluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeu 121
Db 28140 GACCAGAGCATTCATTACTACGAGCAGGTTCCTTCAAAATTATAGGGTGGTACCTG 28199
QY 122 AsnAspProSerGlyPheThrLeuMetGly----- 131
Db 28200 GTAAACCCCGCGCTACGGGTTTCAGGAATAGCTACGGCGCAGCAGGTGAACCTTGCA 28259
QY 131 ----- 131
Db 28260 GGTGGGATCTAGTGATATTTTCTTGAGCGCTGGCATGATCAAGGTGGCGCCACC 28319
QY 132 SerArgIleSerAsnAsnProAsnIleLysLysGluThr---LeuGluProIleGlnLeu 150
Db 28320 AGTGAATGATGAATGAACACCAATCTTGACTCCAAACGCAAGATATTTCCCATCGACTC 28379
QY 151 AspPheAsnAsp----- 154
Db 28380 GATCCAGCAATACCAACTCCTATATATTTTCGAGGAGTCAACCTCAOGGGACCGCG 28439
QY 154 ----- 154
Db 28440 ACCGTTCTCGCACTGTCTCTCCAGCAACAGCACCATCAGCCCAACATATATACGA 28499
QY 155 -----ProThrValAlaLysSer 160
Db 28500 AAAACCGCAGAGACAGCTGGAGTGCTACTACACGGCAAAAGAACCATGGCAGACGGT 28559
QY 161 ProAlaLysThrSerThrAla-----LeuAsnAlaVal 171
Db 28560 ACAGCTGGCAGTTCAAGCTGCTTCCATCAGCGGGAGAGGTAGATAGATAGTCCGCA 28619
QY 172 ValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn---ProTyrPheAla 190
Db 28620 GGAACCTCGTTGACACAAAGCTAGCCACGACCCCGAAAGCAAAATTTTGATCCCTTTGCC 28679
QY 191 LeuLeuGluSerTyrLysGlyAsnGly-----ThrProIleSerThrSer----- 206
Db 28680 ACCAGTATAGCTGGGACGACACCGGAGCCACCAACCGGTGTGCACGGGAGTACAGCG 28739
QY 206 ----- 206

Db 28740 GGAAGCCGTCGTGCTCAATAGACAAGAGCCTGGGGAACAAACAAAGCATCTGCACCTTT 28799

Qy 207 -----AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln 220

Db 28800 GACTACAGGATCCCAAGACCTTACCTATTCGCGCTCGGTACAAATTTTGACAGCTT 28859

Qy 221 GlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerThrGlySerLys 240

Db 28860 GGTAAACCGCATTAATACGACATATTC-----CGCAAAACAGACGACAC 28907

Qy 241 ThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaIaSerGlyThr 260

Db 28908 ACCTGGGACACTACTACTACTCGAA---AAGGAGGATGGCTCGCTCATCTCGCAGAC 28964

Qy 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280

Db 28965 CCCAACGCCATTTCCCTTACCAG-----TTACGCTTTGGCTCTGAGGCCAA----- 29012

Qy 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300

Db 29013 -----CCGCTTGACTCAGC----- 29027

Qy 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320

Db 29028 -----GGTCTCCCC----- 29036

Qy 321 ProLeuThrLeuAspPheGlyIleLysSerGlnAsnMetTrpAlaGlyAlaProAla 340

Db 29037 -----CTTGAC-----GGCCCCCAACA 29054

Qy 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360

Db 29055 GCAACCCAGGAGGCCATCGATTGGCGCAACCGCTCTGAGCCCAACGCCCTATC---ACC 29111

Qy 361 SerSerGlyAsnSerThrAlaAsnGlySerSerSerThrArgArgTyrSerGlnAsn 380

Db 29112 GTTACCTTTGATACCCAGCTTTAAAGTAGCTTCAAGTTATCTCTCAAGAGCAGAAC 29171

Qy 381 GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly 400

Db 29172 GGCTATGGCGGAGGAAACCTGAGCGGAGTAAATATCTGTCGCCAGCGCAATGTTATTC 29231

Qy 401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420

Db 29232 GCATACTCCAAACCGTAAACCCGTAACATTGCGGCAATCTCTGCGTAATTTCCAAAT 29291

Qy 421 GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440

Db 29292 CCTCGGAGCTGACCATGGAGGTAATACTTTTCTGCCAGCGGCACTCAGGAGCA 29351

Qy 441 ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460

Db 29352 CCACGAGTTGCTCTACCCGGA---CCTGAGCTGGTAAAGTTTTCACAAATCTCGCTGGAA 29408

Qy 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleLeuGlnArgGlyPhe 480

Db 29409 CAATCCATATGTCATATGATGTCGAGTTTGTCCGATGATCACCATTCAAGCTGGCTAC 29468

Qy 481 GlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500

Db 29469 CAGGCAAACTCTAAATATCATCACCCGCTTACGACAACTACTTGGCGAAATTAACCTT 29528

Qy 501 LysArg 502

Db 29529 AAACGC 29534

RESULT 9

AE013853/c 12559 bp DNA linear BCT 26-JUL-2002

LOCUS AE013853

DEFINITION Versinia pestis KIM section 253 of 415 of the complete genome.

ACCESSION AE013853 AE009952

VERSION AE013853.1 GI:21959364

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 12559)

Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,

Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,

Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,

Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,

Blattner, F. R. and Perry, R. D.

Genome Sequence of *Yersinia pestis* KIM

J. Bacteriol. 184 (16), 4601-4611 (2002)

12142430

2 (bases 1 to 12559)

Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,

Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,

Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,

Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,

Blattner, F. R. and Perry, R. D.

Direct Submission

Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445

Henry Mall, Madison, WI 53706, USA

Location/Qualifiers

1. 12559

/organism="Versinia pestis KIM"

/mol_type="genomic DNA"

/strain="KIM"

/db_xref="taxon:187410"

complement (87..2153)

/gene="y2497"

complement (87..2153)

/gene="y2497"

/functions="putative"

/note="residues 1 to 688 of 688 are 48.83 pct identical to

residues 1 to 684 of 684 from GenPept : >dbj|BAB53429.1|

(AP03011) probable oxidoreductase [Mesorhizobium loti]"

/codon_start=1

/trans_table=11

/product="putative oxidoreductase"

/protein_id="AA086053.1"

/db_xref="GI:21959365"

/translation="MKSLSWDADAECVERLGLLGLPELGLRTYSARLLGSNPKVL

HGGNTSVKLNMDLFGNPMQVCMICKSGQDLATITAGHPAVRLAPLOQLKVLDDHLS

EDMVNLTLLCRAPTSVEALLHAFVDRPVIDTHAVAAVIAADQPDQDLRCRY

IYGERVWIPVMSGFALSVAERAFRAYPDAGMLLKGHLCTFGGSASAKYERMIT

FVTLAEPIESRAIYSPVTPPAKPNLADNAHIAFYLAARAFARSASISEKKWTF

DIRDPKSLAFSSGLGSDIAQGVATPDHVIKMGPPVLDADPLNLDGMAAQOT

KIAAFERYHYFHRNNARFERKIALDPLRAIVIFYLGVGAGRTAEADAANADLI

ESLWVDVDTATVCGDEYPIGEADFMEYSLQEKIDKASVPLQCGQIVVVGAGGG

ICAAIAKEFSLGLAEALVDIDSEAKNVAOLGPHALACQDVETASVQAAPEMIA

TKFGVDIVVSNAGIALSGAIELPATLRTSFVNFFAHQVAAQAVSIKKQGIIG

VLLNISKQALNPGINFGAYCTSKAALLSVROVALBQGDGIRVAVNADRISGLL

DDEMLSLARARGLSSEKYMAGNLLGQEVTAQDVAKAFVVSAMLDKXGTVITVDGNG

VAAMLR"

1395..1619

/gene="y2496"

1395..1619

/gene="y2496"

/function="unknown"

/codon_start=1

/trans_table=11

/product="hypothetical"

/protein_id="AA086054.1"

/db_xref="GI:21959366"

/translation="WCHVSQIVGLCGGRYHWRGIYGPAPFNKIFSQGDKRNHSLIRGF

GTLTKRTQAMLTQSHTLIRVRPEGTDFRE"

2371..3372

/gene="y2498"

2371..3372

/gene="y2498"

/function="putative regulator"

/note="residues 25 to 327 of 333 are 41.91 pct identical

to residues 28 to 330 of 694 from GenPept :

gene

CDS

gene

CDS

gene

CDS


```

gene
complement(4520. .6184)
/gene="flgk"
/note="synonym: y2501"
complement(4520. .6184)
/gene="flgk"
/function="structural component; surface structures"
/note="residues 3 to 552 of 554 are 57.60 pct identical to
residues 2 to 546 of 547 from E. coli K12 ; B1082;
residues 3 to 552 of 554 are 57.97 pct identical to
residues 2 to 546 of 547 from GenPept ;
>gb|AA055828.1|AE005318.5 (AE005318) flagellar
biosynthesis, hook-filament function protein 1
[Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=1
/product="flagellar biosynthesis protein, hook-filament
function protein 1"
/protein_id="AA086058.1"
/db_xref="GI:21959370"
translation="MNSLMTAMVSGLNAAQYALSTVSNNTITPQVAGYNRQNTVFAQ
NGGTTTSGAGTFCGVNVTGVNRENAFTIQLRASQTCQSSGLATYYQISQIINLISN
ASNNLSITQDFFNLQNLVNSADDDAAKRTVLKAEGLVNFQFNKLYRDMDDGVN
OKITDSATQIINNAEOTAKLNDQITLRLGSSGSEPNALLPQDOLVTELNQIMAVT

```

12264	Db	GGGATTACTCAAGATTTTAAATGACGGTACGGCCACACCACTAAATCGTCGTTGGATCTG	12205
90	QY	AlaileGlyGlyLysGlyPheGlnValThrLeuGluAspLys-----ValHis	106
12204	Db	GCTATCAGCCAGAACGGTTTTTTCGGT-----ATGCAGACAGCAGCGCGGCAATTAC	12151
107	QY	TyThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGly	126
12150	Db	TATGCCCGTAACGCCAGTTTAAAGTGGATGAAACCCGTAATATCGTCAATATGCAAGGT	12091
127	QY	PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu	146
12090	Db	CTAAATCTACGGGTTACCCGGCAACAGGGACACCAACCGGTACACAGGGCGCTAAC	12031
147	QY	ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr	166
12030	Db	CCGGTCCCCTGCTATTCCGAGGATATGATCTCC-----GCCAAGCGCAGCAGC	11980
167	QY	AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn	186
11979	Db	TCCGGCAATATGTTGGCCAACTGACCTACCCATGATGTTATCCGGAGAGGACCTCA	11920
187	QY	ProTyThrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSer	206
11919	Db	CCITTT-----GATCCCGATAACCAGAT	11896
207	QY	AsnTyThrSerTyRAlaGlnProMetArgValTyAspGlnGlnGlyAsnSerHisAspIle	226
11895	Db	ACCTACAGCTTTGTCAATAACATGACGACCTTTGATAGCTTGGGTAAACCGCATGAATC	11836
227	QY	ThrVal---TyThrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyLeu	245
11835	Db	AATGTCCTTTATGTTAAACGTCTGAAGATGCGCGGATACCTTGGGACGCTAC	11776
246	QY	ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu	265
11775	Db	ACCAGGAGCAGTAGCGCCAACTGACGACCCAGCAGACCCGACCGATCTCTGCTGCC	11716
266	QY	LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe	285
11715	Db	GCTAAACGTGGTTTCGATGTTTTTATAGTAGTACCGTGCCTTGAAACCGTGCACAA	11656
286	QY	ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn	305
11655	Db	ACA--AATGCAACCAAGTACCAACAGACTTCAGTTTACGATCCCGATGGCGGTGTTAAT	11599
306	QY	GlyLeuPro-----GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThr	323
11598	Db	GGTGGCCCGCAAAAGTTTCCATCTGAATGTGCGGGCAGT-----	11557
324	QY	LeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla	343
11556	Db	-----AAACAGCAAAAT-----	11545
344	QY	AlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGly	363
11545	Db	-----	11545
364	QY	AsnSerThrAlaArgAsnGlySerSerSerThrArgAgtyrSerGlnAspGlyTyPro	383
11544	Db	-----ACCGTGCTCAGCAGCATTTGTGCAGAACACAGCAGCTGCTACGCC	11500
384	QY	GlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTySer	403
11499	Db	GCGGTGAAATTTACTGGGTTCCAGATCAACAGTGAATGTTCTGTGTGGGACTTATTC	11440
404	QY	AsnSerGlnValValAspPheTyRAsnIleProLeuAlaArgPheThrSerGluAspGly	423
11439	Db	ACCAGCAGACCAATTGCTCGGCCAAATCGTCATGGTTAACTTCTCAACCCAGAAAGG	11380
424	QY	LeuArgArgGluGlyAsnAsnHisTyRSerAlaThrLeuAspSerGlyGlyProGluPhe	443

/protein_id="AA561568.1"
/db_xref="GI:45436012"
/translation="MSGKCKSTRNOKGPMRDEKSLAHTWNCCKYHIVFAPKVRRO
VYREBRAIGSLRLKCEKNVNLAECCVDHLMLEIPKMSVSGMGYLGKGS
SLMYPQGLDFKRNRFWCRGYVDVVGKNTARIQBYIKHQLEDKMGQSLIPY
PSPFFTRK"
gene complement(1183..2274)
/genes="ompA1"
/locus_tag="YP1326"
complement(1183..2274)
/genes="ompA1"
/locus_tag="YP1326"
/notes="highly similar to Yersinia pestis CO92 ompA and
Yersinia pestis KIM ompA. Contains OmpA-like transmembrane
domain (Interpro|IPR000498). Contains MotY protein
(Interpro|IPR01035, (GO:0001539), (GO:0009288)). Contains
OmpA outer membrane protein (Interpro|IPR002368,
(GO:0016021)). Contains Bacterial outer membrane protein
(Interpro|IPR006664). Contains OmpA/MotB domain
(Interpro|IPR006665)."
/codon_start=1
/transl_table=1
/product="putative outer membrane porin A protein"
/protein_id="AA561569.1"
/db_xref="GI:45436013"
/translation="MPTMIMRKKKTAIALAVLGVPAVAQAPKDNWTYGGKL
GMSQYQTSINNDGTHDQKAGAFGYQANQYLGFMGYDMLGRMPYKGDING
AFKAQGVQAALSLVQADLDVYTRGLGLVWRDAKGSFDGGLDRASGHDGVSPLV
ALCAEAWATWATMEYQWNNIGDRETVGARPDNGLSVGVSRYFGQEDAAPIVA
PTPAPAPIVDTRFKLSVDLFGFNKANLKPEQQAOLDQYLAQLSSIDPKGSSVVLG
FADRIQAPNALSORRDSVRDVLVSKGIPADKITARGEQANPVTGNTCDNVKPR
AALIEADRRVEIEVKYKEVTPQQA"
gene complement(2602..3108)
/genes="sula"
/locus_tag="YP1327"
complement(2602..3108)
/genes="sula"
/locus_tag="YP1327"
/notes="highly similar to Yersinia pestis CO92 sula and
Yersinia pestis KIM sula. Contains Cell division inhibitor
Sula (Interpro|IPR004596, (GO:0019866))"
/codon_start=1
/transl_table=1
/product="putative cell division inhibitor"
/protein_id="AA561570.1"
/db_xref="GI:45436014"
/translation="MRTQSLKPHYANYHSLTNDSTRVDAPDTSGLISEFVYSNQP
VYQLLPLQLQSKSRWLLTLPQKLSRWLQSGPLINKVQLQINPLSTVEA
MEKALLTGNYSVVLGWLPELTEDDRIRLRLAALGNAYGVNRPNDTKVSGQCANTL
KIHSYLHV"
gene 3250..3972
/genes="tfoX"
/locus_tag="YP1328"
3250..3972
/genes="tfoX"
/locus_tag="YP1328"
/notes="highly similar to Yersinia pestis CO92 YP01437 and
Yersinia pestis KIM Y2733. Contains TfoX, N-terminal
(Interpro|IPR007076). Contains TfoX, C-terminal
(Interpro|IPR007077)."
/codon_start=1
/transl_table=1
/product="regulator of competence-specific genes"
/protein_id="AA561571.1"
/db_xref="GI:45436015"
/translation="MRIFALSCSSYYCYKGTGLVIRFTKGVVNMKMSKRIAQQRV
FGFLQQTIRKSGGGYGLLVNSIMFVSEGEDILRANONIEDLFRASGMALVAKR
GLPVLRYRVDTSIWNLDGLRLHFVRQAYEGAKIDVSKGPPVRLKDLPLNSASLE
RLWVKYIKNATRLIEGAKCCYLLKRLARRSIGINILLALAGAISGYHYAALPLMVR
SELIWFPEKHHPAKAPLYETA"
misc_feature 3826..3885
/genes="tfoX"
/locus_tag="YP1328"

/note="probable transmembrane helix predicted by TMHMM2.0"
complement(4074..6212)
/locus_tag="YP1329"
complement(4074..6212)
/locus_tag="YP1329"
/notes="highly similar to Yersinia pestis CO92 YP01438 and
Yersinia pestis KIM Y2732. Contains Sugar transporter
superfamily (Interpro|IPR005829, (GO:0016020))"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="AA561572.1"
/db_xref="GI:45436016"
/translation="MFTFVDDLRRYAYNSLLHARIFIALVGVTAVPWVIEQKLI
PLTUGVAAALTDDELTRLRNLFTILICFFVASSIIELLFPYFFALGLTSLIC
GFTLGLQARYATIAFCALLIAYITMLGASMTPTWYQQPLLISGAVWNVLTLSFGH
LIFIRPLQONLARYCYQLAHLEAKANLFDPIEDINQPLIDVAVNVTLSATLNG
AKASLVRLLKRGQRCTRTLRVYFVAQDIHERASSSHFYQVLRKFRYSVDLFRF
CELLSMOAKCNHLSQILMHQYQHPDPRFAPTFUDAARELAHEHNTQCVKAL
SYLLKNRAIDAQLAGISEQVLAQDPTPDNHLSDDEITGMSDIKLIISHLFQPSAL
FRHAIKSIYVLCVGIAFTIQTILRHGWILLTSLFVCPQYSATRRRLAIRLIGTLAG
ILVGLPIYVFSEIGQTLIVISGVLPFAFNVQVYAHATMTILLVLLCFNLGEGF
EVAAPRVDTLLGCAIAWAVSTIPDMKPRQLPTVVKHTFNANCYRLIDAILAQYHQG
KDNLSLYVARRDAHNCDAELASVTSNMAEPKNDKETQEAARLLCLNHTMISYISA
LDGAREKLTNSTTLLNLDACIYIEGTLAKEQPDNARMASALAELSNRLQCHPEPES
RDLVLAQIGLLLELLPELSLNKRISQPA"
misc_feature complement(4611..4870)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(4689..4757)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(4827..4895)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5718..5786)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5796..5864)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5883..5942)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(5952..6011)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(6102..6170)
/locus_tag="YP1329"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(6242..6688)
/locus_tag="YP1330"
complement(6242..6688)
/locus_tag="YP1330"
/note="highly similar to Yersinia pestis CO92 YP01439 and
Yersinia pestis KIM Y2731. Contains Protein of unknown
function DUF307 (Interpro|IPR005185)."
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="AA561573.1"
/db_xref="GI:45436017"
/translation="MRTVLNVNLFVGLGFTTGLMLATLVSVILITFLPLRSCWEI
TKLSPLPYGNEAHVNLVPEKSNALFTAGGALINIVMFIIFGWLCLSHITAGIVQC
MTIIGIPVGINFKIAIALMPVGRVVSVEVAQARVANARQHR"
misc_feature complement(6311..6379)
/locus_tag="YP1330"
/note="probable transmembrane helix predicted by TMHMM2.0"
complement(6413..6481)
/locus_tag="YP1330"

Alignment Scores:
Pred. No.: 1.59e-24 Length: 290155

Score:	557.50	Matches:	145
Percent Similarity:	43.98%	Conservative:	74
Best Local Similarity:	29.12%	Mismatches:	196
Query Match:	21.56%	Indels:	83
DB:	1	Gaps:	11
US-10-009-823A-1 (1-502) x AE017132 (1-290155)			
Qy	10	ThrGlyMetThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29	
Db	290028	AGCGTATGAACGAGCAGCAGTACAGTATGATCGTAACTAACAATATCGCCAACTCG 289969	
Qy	30	AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49	
Db	289968	GCAACCTCAGGTTTAAAGCAGGTTCTCTTTCCTGATATGTTTTCG- 289918	
Qy	50	AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69	
Db	289917	-----GGTTCT-----CAGACCGCATGGCGGTAAAGTTCGG 289885	
Qy	70	SerValArgThrIlePheThrGlnGlnAlaPheGluProGlyAsnSerValThrAspLeu 89	
Db	289884	GGGATTACTCAAGATTTTAAATGACGGTACCGCCACCACTAATCGCTTTGGATCTG 289825	
Qy	90	AlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLys-----ValHis 106	
Db	289824	GCTATCAGCCAGACGGTTTTCCTG-----ATGCAAGACAGCAGCGCGGCATTTAC 289771	
Qy	107	TyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGly 126	
Db	289770	TATCGCGTAAACGGCCAGTTTAACTCGATGCAAAACCGTAATATCGTCAATATGCAAGT 289711	
Qy	127	PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu 146	
Db	289710	CTAAATCTAACGGTTACCGGCAACAGGACACCAACCGGTACACAGGCGGCTAAC 289651	
Qy	147	ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166	
Db	289650	CCGGTCCCGCTGTCTATTCCGACGATATGATCTCC-----GCCAAGGCGGACGACC 289600	
Qy	167	AlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186	
Db	289599	TCCGGCAATATGGTGGCAACCTGACCTTACCATGATTTATCGCGAAGCGACCTCA 289540	
Qy	187	ProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProIleSerThrSer 206	
Db	289539	CCTTT-----GATCCCGTATACCCAGAT 289516	
Qy	207	AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226	
Db	289515	ACCTACAGCTTTGTCAATAACATGACGACCTTTGATAGCTTGGTAAACCGCCATGAATC 289456	
Qy	227	ThrVal---TyrPheAspGlyAlaProSerSerThrLysSerLysThrPheGluTyrLeu 245	
Db	289455	AATGCTTTTATGTTAAACGTCTGAAGATCGCACCGATCGGAATCTTGGACGCTTAC 289396	
Qy	246	ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265	
Db	289395	ACCAGGACAGTAGCGCAAGTACCGACCGACCGACCGACCGATCTGCTGCTGCC 289336	
Qy	266	LeuMetSerGlyThrMetThrPheSerSerAsnGlyGlnLeuLysAsnMetThrAlaPhe 285	
Db	289335	GCTAAACGTGTTGATGTTTGTAGTAAACCGTCCCTGAAACCGTGAACACCGT 289276	
Qy	286	ThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsn 305	
Db	289275	ACA---AATGCAACAGTACCACAGACTTCACGTTTACGATCCCGATGGCGGTGTTAT 289219	
Qy	306	GlyLeuPro-----GlnPheSerAlaAsnPheValGlyAlaGlyLeuGlnProLeuThr 323	
Db	289218	GGTGGCGCCGACCAAGTTCGCACTGAATGTTGCGGGCAGT----- 289177	
Qy	324	LeuAspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAlaSerAlaAla 343	
Db	289176	-----AAACAGCAAAAT----- 289165	
Qy	344	AlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGly 363	
Db	289165	----- 289165	
Qy	364	AsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyrPro 383	
Db	289164	-----ACCGTCTGACACCATTTGTCACAGAACCCAGACTGGGTAGCC 289120	
Qy	384	GlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSer 403	
Db	289119	CGCGTGAATTAATCTAGCTTCCAGATCAACAGTATGATGTTCTGTTGGTAGGACTATTC 289060	
Qy	404	AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423	
Db	289059	AACAGCAGACCAATTCCTCGGCCAAATCGTCACTGTTAACTTCTTAACCCAGAGGG 289000	
Qy	424	LeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443	
Db	288999	CTGTCTCCGAGGCGATACGTGGAAAGAGACGAGTCTTCGGTAAACCCAGCGTC 288940	
Qy	444	GlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn 463	
Db	288939	GGGACTCGCGGTAGCGGTGCTCGGCACGTTGACAGTGTGCGCTGGAATCTCTAAC 288880	
Qy	464	ValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsn 483	
Db	288879	GTGGATTGATGAAGATGTTGTAACATGATTGGCGCAACGTAATATCATCTCCCAAC 288820	
Qy	484	SerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501	
Db	288819	GGCGAGACCATCAAAACCAAGATCAGATCTTACAACTCTGTTAGCTGCGT 288766	
RESULT 11			
AJ414150			
Yersinia pestis strain CO92 complete genome; segment 10/20.			
DEFINITION			
AJ414150 AL590842			
VERSION			
AJ414150.1 GI:15979723			
KEYWORDS			
Yersinia pestis CO92			
SOURCE			
Yersinia pestis CO92			
ORGANISM			
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
Enterobacteriaceae; Yersinia.			
REFERENCE			
1 (bases 1 to 335050)			
AUTHORS			
Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,			
Prentice,M.B., Sebaihia,M., James,K.D., Churcher,C., Mungall,K.L.,			
Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,			
Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,			
Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,			
Karllyshv,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K.,			
Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrall,B.G.			
Genome sequence of Yersinia pestis, the causative agent of plague			
Nature 413 (6855), 523-527 (2001)			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
21470413			
11586360			
REFERENCE			
2 (bases 1 to 335050)			
AUTHORS			
Parkhill,J.			
Direct Submission			
TITLE			
JOURNAL			
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia			
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,			
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk			
Notes:			
Details of Y. pestis sequencing at the Sanger Centre are available			
on the World Wide Web.			
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).			
FEATURES			
source			
1. 335050			
Location/Qualifiers			
/organism="Yersinia pestis CO92"			
/mol_type="genomic DNA"			
/strain="CO92"			

```

/db xref="taxon:214092"
/notes="biovar: Orientalis"
147. .536
/gene="YPO1693"
CDS
147. .536
/gene="YPO1693"
/notes="Similar to Streptomyces coelicolor hypothetical
protein SC664.19C TR:086782 (EMBL:AL031317) (190 aa) fasta
scores: E(): 6.5e-10, 41.3% id in 104 aa, and to Bacillus
halodurans BH0898 protein TR:Q9KEF4 (EMBL:AP001510) (134
aa) fasta scores: E(): 0.034, 30.5% id in 95 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC90513.1"
/db_xref="GI:15979724"
/db_xref="EMBL:Q8ZFL0"
/translation="WATPKQYEDVERWAFGTEQVADDELALVLNGTKTKTCAALDD
EGVPQAGDIFVVGNGRNPACAVELTEVELKTFDQVDEHAALAEGBDSTLAYWRKTO
QRFFERYDMFSDDMLCMCKFKVLETF"
complement (646. .885)
/gene="YPO1694"
CDS
complement (646. .885)
/notes="YPO1694"
/notes="Similar to Escherichia coli hypothetical protein
YebV SW:YEBV ECOLI (P76274) (78 aa) fasta scores: E():
4.1e-19, 57.9% id in 76 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC90514.1"
/db_xref="GI:15979725"
/db_xref="EMBL:Q8ZFK9"
/translation="WKSIVKIROPEVDDAELSSOTKGEHTLSIPCKSDPDLQMLDQ
WENISIPALLGKDTLLYKQYDQHDANWKVT"
complement (1093. .2082)
/gene="YPO1695"
CDS
complement (1093. .2082)
/notes="YPO1695"
/notes="Similar to Pseudomonas aeruginosa hypothetical
protein TR:AG08040 (EMBL:AE004879) (315 aa) fasta scores:
E(): 0, 41.5% id in 299 aa, and to Pseudomonas fluorescens
hypothetical 23.6 kDa protein TR:Q9KIM7 (EMBL:AF228767)
(231 aa) fasta scores: E(): 1.2e-13, 29.9% id in 231 aa"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAC90515.1"
/db_xref="GI:15979726"
/db_xref="EMBL:Q8ZFK8"
/translation="MLSITREITTTINTPKYLSWRTLPIFLWVGNLTACTFSPT
VILPPNSFNVTLTESGSGTAMLCGTGVSLLASPVLNATVASTTNTLKNAGD
VIPYTIYDAGYNNAIAGNIVNAGNSLLIVTLFOISANLPIVIRGNNAINNV
SAGTYDNLNKNYDLCALISFLSCLGGRFSGTATSTVTKLVNSDCVNSAPNV
FGSALITQAPNQONISATKNAFTFTYITNGNVPVAPRMLKLTSTDYLOYQLY
QHGVIYFWNTKRTATGTQQLIPYNAINVPPQAKTVGYQDNVSVLEY"
complement (1663. .1728)
/gene="YPO1695"
misc_feature
/notes="one of 3 probable transmembrane helices predicted
for YPO1695 by TMHMM2.0"
complement (1819. .1884)
/gene="YPO1695"
misc_feature
/notes="one of 3 probable transmembrane helices predicted
for YPO1695 by TMHMM2.0"
complement (1945. .2010)
/gene="YPO1695"
misc_feature
/notes="one of 3 probable transmembrane helices predicted
for YPO1695 by TMHMM2.0"
complement (2288. .4735)
/gene="YPO1696"
CDS
complement (2288. .4735)
/notes="Similar to Escherichia coli outer membrane usher
protein AfaC precursor SW:AFAC ECOLI (P53517) (859 aa)
fasta scores: E(): 1.4e-13, 22.7% id in 631 aa and
Pseudomonas aeruginosa hypothetical protein Pa4652
TR:AG08039 (EMBL:AE004879) (790 aa) fasta scores: E(): 0,
42.9% id in 805 aa"
/codon_start=1
/transl_table=11
/product="probable outer membrane usher protein"
/protein_id="CAC90516.1"
/db_xref="GI:15979727"
/db_xref="GOA:Q8ZFK7"
/db_xref="EMBL:Q8ZFK7"
/translation="MIIRFLPFLSVLIYGSILPLSLITILGSLPPSPSTTTMPD
TLYLELVNDRNFSTVPSYNNRYTYSQSQTITGLPISEPLAEPIDNAGYN
VKYDGENORLLINVPSEWLPKQIEVTEQDQNLGSLGALFNVDIATQGYPSYL
THPSATEQRIEDRFGLLSNTGYRTHFRPSNNTDDAKGYRFDTOQKNDDEHLRY
STGDLITAGPMSAARLGGIQIARHAIERDILTYELPQFSGQAAVPSIVDLYDNF
RTQSANINPGFFVINRAPRINGAGQATIVITDAGIQISVSVFFYASILLKGVDF
SLGGALLRAYAISADYGEWASGVRYGTITPWLITLEGKDAKEMHVIGGGVFRM
GLGLVNLNYSISNTNGAFNVAEPLNTNATPNRLPSLPAASRRGNQRSJGYSYS
NAFNLNAQHIIISDEYSLANTKTPSLSRMTQLTSLGSLGSGYGTGSGSYDVRDA
LGEOTRLINISYSTLLRNSFYALNRELGRKYNQVLMVSIPLGPRGSSISATRT
NDNQIOLNYSRSPNSGGLGNLAVANSTNNNOVQADIVRYSMESRGLYGN
SNYNYVGLTGSILVWRSVYASNMINDAFALVSTNGFSNI PVSYENQIGITNAKG
YLLIPTVASYIQAKFDQDNLNLPADWMLPNVERLALGERSGYLINEPIKEISAVNIR
ITDASGDLPRGSAIYITGNIPISYVGWGNVIEQVAQNNLRILRADNGTCYCSOF
KLKTTEGIQDAGTTVCR"
complement (4923. .5675)
/gene="YPO1697"
CDS
complement (4923. .5675)
/notes="YPO1697"
/notes="Similar to Pseudomonas aeruginosa PA01 probable
pili assembly chaperone TR:AG08038 (EMBL:AE004879) (262
aa) fasta scores: E(): 0, 43.2% id in 243 aa, and to
Escherichia coli chaperone protein CscC precursor
SW:CSCI ECOLI (P53518) (232 aa) fasta scores: E(): 0.0013,
21.9% id in 242 aa"
/codon_start=1
/transl_table=11
/product="putative chaperone protein"
/protein_id="CAC90517.1"
/db_xref="GI:15979728"
/db_xref="EMBL:Q8ZFK5"
/translation="WIKYHLISKLISFILLPIAPLHAASVLIWPIIDPVINDENAT
ALMLEKDSQFTYQIRVLSRWQVNGLDNDAGTALIASPVANIQKQQLIRLVKN
AQVPAKELAYRILIDIEPQEPNDKNPSTQLGLTFQMYRISVPLFVYNGMQIAPE
ATOPTDIAQPTSLYNLRQKQKWLIRNQSEIHARISQVTLNGQTVNNGLLGYLAR
SOMSPFISLAARKGELAAIINSNTKPVITFQ"
complement (5706. .6263)
/gene="YPO1698"
CDS
complement (5706. .6263)
/notes="YPO1698"
/notes="Similar to Myxococcus xanthus protein U precursor
Pru SW:PRU MYXA (P27755) (179 aa) fasta scores: E():
6.7e-05, 25.0% id in 180 aa, and to Pseudomonas aeruginosa
PA01 hypothetical protein TR:AG08037 (EMBL:AE004879) (180
aa) fasta scores: E(): 8e-16, 34.5% id in 174 aa. This CDS
shares 31.9% and 30.8% identity with YPO1699 and YPO1700
respectively"
/codon_start=1
/transl_table=11
/product="putative exported protein"
/protein_id="CAC90518.1"
/db_xref="GI:15979729"
/db_xref="GOA:Q8ZFK5"
/db_xref="EMBL:Q8ZFK5"
/translation="MISLFPCHLSAALLISALSI VFTNAAPSVSPFTVNAVLTVG
CVLGGGATDSNFGISNFGNITSLATNATGNTITICQNGNPSVTLNLSGAN
MTGNISAGRHLLNSSTGEYLLIYIQNSPRTLWGDGNGGATQVIATDGTLOQVILY
AQLFASNTQPTAGVYLDLIVTVTY"
complement (5976. .6002)
/gene="YPO1698"
misc_feature
/notes="PS00221 MIP family signature."

```


Girons, I., Somerville, R.L., Wen, Y.M., Shi, M.H., Chen, Z., Xu, J.G. and Zhao, G.P.
 Unique physiological and pathogenic features of *Leptospira* interogens revealed by whole-genome sequencing
 Nature 422 (6934), 888-893 (2003)
 2 (bases 1 to 10029)
 Ren, S., Fu, G., Jiang, X., Zeng, R., Xiong, H., Lu, L., Lu, G., Jiang, H., Ding, Y., Jia, J., Tu, Y., Gu, W., Cai, Z., Sheng, H., Yin, H., Zhang, Y., Zhu, G., Wang, S., Shen, Y., Qiang, B., Chen, Z., Wen, Y., Xu, J. and Zhao, G.
 Direct Submission
 Submitted (12-MAR-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China
 Updated information will be available at our World Wide Web site (<http://www.chgc.sh.cn/lep/>). Comments to the authors are appreciated.

FEATURES

source

1..10029
 /organism="Leptospira interrogans serovar Lai str. 56601"
 /mol_type="genomic DNA"
 /strain="56601"
 /serovar="Lai"
 /db_xref="taxon:189518"
 /chromosomes="I"
 /note="Strain 56601 is maintained by the Institute of Epidemiology and Microbiology, Chinese Academy of Preventive Medicine"

gene

complement(148..1542)
 /gene="flgE"
 /locus_tag="LA2848"
 complement(148..1542)
 /gene="flgE"
 /locus_tag="LA2848"
 /codon_start=1

CDS

/transl_table=11
 /product="flagellar hook protein"
 /protein_id="AA050047.1"
 /db_xref="GI:24196693"
 /translation="MRSLSYSGVGLKQHVMDVIGNNISNVNTHGKTERVTFQDM
 ISQELRGASPERKIGVNPQVQGLSLIAADKIMTQSLQTTCKNTDVAISGEGRF
 IVKDGKQVPTAGFLDKNGVYVNPANGLVQVNSLDDKNGKYNLSAIEDII
 IFVYSKEPARATSQIDFKSNLSAPAPPDQATQBEITAMINDPDKMRGHTVINT
 PDQGIQRFEYKFRVNDTWKRLSMTDLSQSDVSGTGQDQKQPNTEIEFGF
 TPKGLVYSDGVDSNMGNKAKVFRIPGNPAIQNFDLNLGEAGVNGITQFSSDF
 TTKAVQDGYTGYLESIDNSGITGVFSGVGRQPLARVATAVFNPNAGLDKAGDT
 MFSYNSNGEPNIGAGVQGRGKNAGLLENSVDLSQDTDMIVTQRFQAGSRITIT
 TSDQMTQEVILGLKR"
 complement(1570..2364)
 /gene="flgP"
 /locus_tag="LA2849"
 complement(1570..2364)
 /gene="flgP"
 /locus_tag="LA2849"
 /codon_start=1

gene

/transl_table=11
 /product="flagellar hook assembly protein"
 /protein_id="AA050048.1"
 /db_xref="GI:24196694"
 /translation="MPEATGISQANINRYLEGDRSFNIRNMENLEKEEKNGLKGI
 IRSTVSKLGDQDFLLITQLSSQDPTNPVKQDFIAQMAQFSSLEOMNINSLGQKM
 GNRQSFSLGKLVSGDFVNGENIAGIAGALFDDGKGTFFRVNGRSIDVQIISLD
 PVLKQEAAYNQSSQSPARTAHSEIGTVSPQAPEDISSNAHEMTTKQKSM
 ESSIKNQKEVFGREQFELQEKITSEKKISDWKFPFGKDSNSTE"
 complement(2379..3944)
 /locus_tag="LA2850"
 complement(2379..3944)
 /locus_tag="LA2850"
 /codon_start=1

gene

/transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA050049.1"
 /db_xref="GI:24196695"

CDS

/transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA050049.1"
 /db_xref="GI:24196695"

/translation="MNIISGDLISYSEPKPOQSMESLPMGEVAGKNSFVLMKFLQSGA
 QKGLTEALAIQNTSSKAELQPDEEDPEIIEKKRSINSIEEISKEENFTTQBA
 PNANILPFWLSETKTNESVDPIESVILDELSQSEIKTNSTETLNAPEISSTSELIE
 NFANAEVESTKLLSEVLSKEELNIDTKGSLVSPFNILKEIKARSTQSEIYEKES
 ESFEAKDSOTFRFNEVDKKEKENTKGLSNSFKDQFERAKVEKSELALDSKMKKI
 SRDKTDSYNTAKAEEIKAVVLNQFTENSFKSGSQEOQSPRSDSYSLVKGVTGA
 ANSGREGTFSKEFLPSKESNVLSKDKIQONFONLIRSAVQIENGRTEASIRMPNPK
 DLQMSLAISTDKDVVRKLLVSDTVKQQLVAELANLKQDLKANGLELESILVIEVKE
 REGGSFNADSKNGKDSHSAFAGDMNSDFKHSFYEBEDELSEFENFSEPNDFPKK
 TDKSEKLLDLKV"
 4359..4496
 /locus_tag="LA2851"
 4359..4496
 /locus_tag="LA2851"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050050.1"
 /db_xref="GI:24196696"
 /translation="MLSFPYFWVDDGGGKVRKFLYQKNILFANKKDSFLSEHLKNVC
 P"
 complement(5417..5797)
 /locus_tag="LA2852"
 complement(5417..5797)
 /locus_tag="LA2852"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA050051.1"
 /db_xref="GI:24196697"
 /translation="MDVSYFFGATQYVSIPIFLKEYSVRKPELENGDFOIEABFPFS
 YLTHCKKQSYRFDSEGLIRINDYTADIVGSNAKGARHTTFHQDIEGPPMATKRDVEV
 RLGSWATSIPLSAGELKPLGVYFR"
 complement(5898..6128)
 /locus_tag="LA2853"
 complement(5898..6128)
 /locus_tag="LA2853"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050052.1"
 /db_xref="GI:24196698"
 /translation="MDPFEKLVNRYGGWELWNRDLDFVEFYFQSLSGPLFLKGLGKTF
 NRGKGVFPKREITFLDFPNFKKNIQKW"
 complement(6485..7084)
 /locus_tag="LA2854"
 complement(6485..7084)
 /locus_tag="LA2854"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050053.1"
 /db_xref="GI:24196699"
 /translation="MRKIMILLPCLVFASCEDEKKDQSGITGNSIMDLLLLQEVST
 PPGPGVWMDIPSTINAQVTSKPSFIQFSGSVNHELMKNKCNFSELSVTNL
 PAGLTLNSTGAINGAPTALISANTVTYTSAKLKANNSTPIPTTKTTVTIVFAGSLTC
 NTAGALGCONAALPYSCNSNFCYSTYSSCKAASCGY"
 7397..7531
 /locus_tag="LA2855"
 7397..7531
 /locus_tag="LA2855"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA050054.1"
 /db_xref="GI:24196700"
 /translation="MIFINKITPKKLTFCEIRIFRIHPKMCNLKASLILFTKYRKTKQK
 P"
 complement(7860..8000)
 /locus_tag="LA2856"
 complement(7860..8000)
 /locus_tag="LA2856"
 /codon_start=1

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAN50055.1"
/db_xref="GI:24196701"
/translation="MTLLNDIYCKRIKIEAKINEQCKDVEILENWSRNSRFPRIIL
SF"

```

```

gene      8110.. 9156
/locus_tag="LA2857"
CDS       8110.. 9156
/locus_tag="LA2857"
/codon_start=1
/transl_table=11
/product="predicted hydrolases or acyltransferases,
alpha/beta hydrolase superfamily"
/protein_id="AAN50056.1"
/db_xref="GI:24196702"
/translation="MSCLSKSYEEKRMKILKWLGIITPFLVVTYATPKYEY
KSVPLYSNFDYSYRKLQISRSKVRPGNEELKLYSADKTDPSILYIHGPGASRAEG
EVTDLAKDPKANLYKPKVHALVLPSPYDTPNPFSVIYOFQSWCKDFANLWVGKIR
IGTSMGLISTYLAAYPEKHALVLPSPYDTPNPFSVIYOFQSWCKDFANLWVGKIR
KSTEEKRNPASAFWRDXYLAAVONLSDLREFILCTDPFSKISPIILLFYFKNEKN
QDVSAVSVMNLAKVNGKASPLNKAIVKVEFGNHLVFSKYMKSDDLLILKETFP
IQNVFSLNKLNSH"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 9,05e-26 Length: 10029
Score: 550.00 Matches: 149
Percent Similarity: 43.63% Conservative: 84
Best Local Similarity: 27.90% Mismatches: 199
Query Match: 21.27% Indels: 102
DB: 12 Gaps: 12

US-10-009-823A-1 (1-502) x AE011450 (1-10029)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Dy 1542 ATGATGAGGTCACCTTATTCGGGTGATCCGGACTTAAGAACCCACCGTCCGAATGGAT 1483
Qy 21 ThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrlsGlnGlnValVal 40
Dy 1482 GTCAATCGAACAACATCTCAACAGTAACACTCACGGTTTAAACCCGAGGGTTACG 1423
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
Dy 1422 TTTTCAGGATATGATTTCTCAGGAGCTTAGAGGTCTCCGAGCCTAAAGAGAAATATGGA 1363
Qy 56 SerGlnGlyProAsnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Dy 1362 GGTGTGAATCCTCAACAGTTGTTGGGTCTTTGATCGCTCGATTGATAAATCATG 1303
Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
Dy 1302 ACTCAAGGTTCCCTTCAACACCACTGGTATAAATACCGAGCTTCGGATTTTCAGCGAAGT 1243
Qy 96 PhePheGlnValThrLeuGluAsnLysValHisTyrrArgAlaGlyAsnPheArgPhe 115
Dy 1242 TTTTTCATCGTCAAGGACCGGTGACAAACAATTTTACGAGACCGCGTCTTTAACCTC 1183
Qy 116 ThrGlnAspGly---PheLeuAsnAspProSerGlyPheThrLeuMetGly-----Ser 132
Dy 1182 GATAAAACCGGTTACTACGTAAACCCCTGTAATGTTCTTAAGGTTCAAGGTTGGAACTCT 1123
Qy 133 ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe 152
Dy 1122 AGACTGGATGAAGAAGAAATAATACATCAATTCGGCAGCTTCGATTGAA---GATAT 1066
Qy 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal 172
Dy 1065 ATAATTCAGTATATTTCTAAGAACCTGCAAGGCTACTTCTCAGATAGATTATTAATCC 1006
Qy 173 AsnLeuGlyAspSerThr-----AspLysThrGlnSerGluAlaAsnPro 187

```

```

Db 1005 AATTGTAATCTCTGCTCCGGCGGTTCCGCTGACGCAACCAAGAAAGAAATCACT--- 949
Qy 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheIleSerThrSerAsn 207
Dy 948 -----CGCATGATTAAACGAT-----CCAGATCCTAAGATGAGA 916
Qy 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlnGlnGlnGlnGlnGlnGln 227
Dy 915 AGGGACATGTTACCACGATTATACATTGACGATCAAGGATTCAAGAGAAATTTAAA 856
Qy 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
Dy 855 ATGGAATTTTCAAGAGTTCTGTGATAACACTTGGAGGCTAGATTGACGATGACGAGTTCA 796
Qy 248 MetAsnProSerGluAspGlySerAlaAsaSerGlyThrAspSerAla----- 263
Dy 795 ACTCAACTTCCGTTGATGTTCTGGACCGCGGCAAAACACTCACTCCAGGTAAC 736
Qy 264 -----GlyLeuLeuMetSerGlyThrMetThrPheSerSer----- 275
Dy 735 ACTGAATTAGAAATTTGCTTTTACTCCGGATCGAAACTAGTATACGTTTCTGACGGTGTG 676
Qy 276 -----AsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer--- 290
Dy 675 GATTCGATGAACACGCGTAAAGTTGACGCAAAAGTTTCTTTTCGGATCTCGGAATCCT 616
Qy 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuPro 308
Dy 615 CGGATTCAAAATTTGATCTTAACCTTAGGAGAA---CGAGGAATGGTAAACGGAATCACT 559
Qy 309 GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIle 328
Dy 558 CAATTCCTCTCTGACTTT----- 541
Qy 329 LysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp 348
Dy 541 ----- 541
Qy 349 IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg 368
Dy 541 ----- 541
Qy 369 AsnGlySerSerSerThrArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuVal 388
Dy 540 -----ACTCAAAAGCTGTCAACACAGACCGTTATACGATGGTATTCTAGAA 493
Qy 389 AspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValVal 408
Dy 492 TCTTTTTCGATCGATAAATCTCGAACCGATTACCGGTGTGTTTCGACCGGAGTCGCTCAG 433
Qy 409 AspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGly 428
Dy 432 CCTCTTCGATGAGTTGGACGCTCGGTTTAAATATCCCGCGGTTTAGATAAAGCGGGT 373
Qy 429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThr 448
Dy 372 GATCAATGTTTTCTGATTCCATCACTCGGAGAACCTAACATCGGTGAGCGGAGTC 343
Qy 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
Dy 312 CAAGAGAGAGTAAATCAACGCGAGGACTTTTAGAGATGTCATATGTTGATCTTCCGAT 253
Qy 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Dy 252 CAGTTTACGGATATGATCGTGACTCAGAGAGGTTTTCAAGCTAACTCTAGAACGATTACT 193
Qy 489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Dy 192 ACTTCCGATCAGATGATCAAGAGGTTTGGGACTCAACGCT 151

RESULT 13
AE0117291
LOCUS

```


DEFINITION Leptospira interrogans serovar Copenhageni str. Fioacruz Li-130,
chromosome I, section 5 of 14 of the complete sequence.
ACCESSION AE017291
VERSION AE017291.1 GI:45600119

SOURCE
ORGANISM Leptospira interrogans serovar Copenhageni str. Fioacruz Li-130
Bacteria; Spirochaetes; Spirochaetales; Leptospiaceae; Leptospira.

REFERENCE 1 (bases 1 to 300594)
Nascimento,A.L., Ko,A.I., Martins,E.A., Monteiro-Vitorello,C.B.,
Ho,P.L., Haake,D.A., Verjovski-Almeida,S., Hartskeerl,R.A.,
Marques,M.V., Oliveira,M.C., Menck,C.F., Leite,L.C., Carrier,H.,
Coutinho,L.L., Degraeve,W.M., Dellagostin,O.A., El-Dorriy,H.,
Ferreto-E.S., Ferreto-M.I., Furlan,L.R., Gamberini,M., Giglioti,E.A.,
Goes-Neto,A., Goldman,G.H., Goldman,M.H., Harakava,R.,
Jerônimo,S.M., Junqueira-De-Azevedo,I.L., Kimura,E.T.,
Kuramae,E.E., Lemos,E.G., Lemos,M.V., Marino,C.L., Nunes,L.R., De
Oliveira,R.C., Pereira,G.G., Reis,M.S., Schriefer,A.,
Siqueira,W.J., Sommer,P., Tsai,S.M., Simpson,A.J., Ferro,J.A.,
Camargo,L.E., Kitajima,J.P., Setubal,J.C. and Van Sluys,M.A.
Comparative Genomics of Two Leptospira Interrogans Serovars Reveals
Novel Insights into Physiology and Pathogenesis
J. Bacteriol. 186 (7), 2164-2172 (2004)

TITLE Nascimento,A.L., Ko,A.I., Martins,E.A., Monteiro-Vitorello,C.B.,
Ho,P.L., Haake,D.A., Verjovski-Almeida,S., Hartskeerl,R.A.,
Marques,M.V., Oliveira,M.C., Menck,C.F., Leite,L.C., Carrier,H.,
Coutinho,L.L., Degraeve,W.M., Dellagostin,O.A., El-Dorriy,H.,
Ferreto-E.S., Ferreto-M.I., Furlan,L.R., Gamberini,M., Giglioti,E.A.,
Goes-Neto,A., Goldman,G.H., Goldman,M.H., Harakava,R.,
Jerônimo,S.M., Junqueira-De-Azevedo,I.L., Kimura,E.T.,
Kuramae,E.E., Lemos,E.G., Lemos,M.V., Marino,C.L., Nunes,L.R., De
Oliveira,R.C., Pereira,G.G., Reis,M.S., Schriefer,A.,
Siqueira,W.J., Sommer,P., Tsai,S.M., Simpson,A.J., Ferro,J.A.,
Camargo,L.E., Kitajima,J.P., Setubal,J.C. and Van Sluys,M.A.
Direct Submission
Submitted (29-FEB-2004) Laboratório de Bioinformática/Instituto de
Computação, Universidade Estadual de Campinas, Avenida Albert
Einstein 1251 Box 6176, Campinas, SP 13084-971, Brasil
Location/Qualifiers
1..300594
/organism="Leptospira interrogans serovar Copenhageni str."
Fioacruz Li-130"
/mol_type="genomic DNA"
/strain="Fioacruz Li-130"
/serovat="copenhageni"
/db_xref="taxon:267671"
/chromosome="I"
complement(186)..1310)
/locus_tag="LIC10995"
complement(186)..1310)
/locus_tag="LIC10995"

Gene Gene complement(186)..1310)
CDS CDS complement(186)..1310)
located using Blastx/Glimmer"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA06906.1"
/db_xref="GI:45600120"

/translation="MKIQQLQKFRVGLTLCPLIFLFGWNSYAKIELTYDLII
GNFYSLTKTVYTINVAVKYVRDGPSPILFKPSVLVFVGDNSTLTPELAKEILL
KGKAQNYIMDLPHGASTMAGPSEYPANYSQLSVTNVADARALLGRMTSTEKKLI
QTIVGHSGLVITQIDLFHRNGNLSSGFIENTLLIGSDIPISLPWPYGDDPMWD
PKSAKSFWNNRFLIKLSLDPPDYAFEGFPVKTPDDFTINTKIYANGVNVTGAPTE
QUKTMSDLPEYYAANAIVGLDTATTDVPIRSVAQNLRWNFNKLVMWMKDGVFFSQ
SETQGLAQYLKELGNALITSDEATHSNPSFKPSLEFLSF"

gene gene
CDS CDS
/locus_tag="LIC10996"
2314..3207
/locus_tag="LIC10996"
/note="PAL domain; identified by sequence similarity;
putative; ORF located using Blastx/Glimmer"

gene
 /db_xref="GI:45600125"
 /translations="MNNRIIESKYTYLENQKYLFLSYSETSDSKNVILFVHGYPDTH
 KTWDLQVNLKERTLGAIDRSGRSKSPSEQSEYNYTMIISDLLEAIRFLSKKKV
 ILVCHDGAALGLWLFSDPYSKYVSFTAISGPHWLAGKRIIDDLFSFDLSNWSKV
 IDGERSWYWFQICVPLPELIWQNGESTYKWMIDWGGVPKSKSLRKVNKNDIYNST
 IAPINLRELLFKTYVSSSHKIKVPQLIVPQKDFIVLPEVYENSIDYVEKLEIHKI
 PSNVHREQPDVLVLELIQKVFVKSI"
 8373..8744

gene
 /locus_tag="LIC11001"
 /note="transposase; IS1533; this region contains an
 authentic frameshift, and is not the result of a
 sequencing artifact"
 8746..8954

CDS
 /locus_tag="LIC11002"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="transposase, IS1533"
 /protein_id="AA569612.1"
 /db_xref="GI:45600126"
 /translations="MNNLHPENILQETFTFYYSRVVQENLKAMDQKIQIEASEP
 FKERVGLRCFRGVDTLT"
 complement(9714..11381)

gene
 /gene="lipL71"
 /locus_tag="LIC11003"
 complement(9714..11381)
 /locus_tag="lipL71"
 /locus_tag="LIC11003"
 /note="lipoprotein; identified by sequence similarity;
 putative; ORF located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="lipL71"

CDS
 /protein_id="AA569613.1"
 /db_xref="GI:45600127"
 /translations="MKSOTKLSFIILLAGFFISCGAELPIEELSDAKNSITRAKS
 AGAEKAPSLERAKNLTATOHOKASENLTETKSAIYARAKALDASEKSPSSVDD
 ARKSSSISSAEAYASQALASPYNTSVOLKEGSLRETDRTDLYSKPKSGDDAK
 LRMLAFDQYERAKIADKAADESKVLSLQKQLIDSFADIKNLNADKYAE
 KQPEVSTNRLSDSKSKEGKEGSEIDDIRKSGELVAKNIKILYAEKQKELA
 KQVSAATTLAFLDRKNSRDFQVSYORAEENLKAAEESRVAEDFLYSEKVEDS
 ISRSEAIRLSRLVQATLAEIRKATTDKIDAGRTKTEGNKTNKQSTEGKNS
 SSIKIGDELPEGWKRYVVRKVPADCLWRIAKDRHYGTSLKWRITYEARNKIKNPN
 LIYKQVLLIPPRKPTRLDKVESAPSRKKKPAEVEALEENRKPETHSSSEDSEAGE
 SNKKKARATPPSTDSEEGGESSENESSPEENGEEENPENLQ"
 complement(11429..11764)
 /locus_tag="LIC11004"
 complement(11429..11764)
 /locus_tag="LIC11004"

gene
 /note="identified by sequence similarity; putative; ORF"

Alignment Scores:
 Pred. No.: 5,09e-24 Length: 300594
 Score: 550.00 Matches: 149
 Percent Similarity: 43.63% Conservative: 84
 Best Local Similarity: 27.90% Mismatches: 199
 Query Match: 21.27% Indels: 102
 DB: 12 Gaps: 12

US-10-009-823A-1 (1-502) x AE017291 (1-300594)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 251203 ATGATGAGGTCACCTATTCGGTGTATCCGACTTAAGAACCAACGAGTCCGAATGGAT 251262
 Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 251263 GTCAATCGGAACACATCTCCAGCTAAACATCAGCGTTTAAACCGACGGGTACG 251322
 Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55

Db	251323	TTTCAGGATATGATTCTTCAGAGCTCAGAGCGCTCCAGACCTAAAGAGAATATTGGA	251382
Qy	56	SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe	75
Db	251393	GGTGGAATCCTCAACAAGTTGGTTCTTGTATCGCTGCATGATAAATCATG	251442
Qy	76	ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyGly	95
Db	251443	ACTCAGGTTCCCTCAACACCACTGGTAAATACGAGCTTGGATTTCAGGCAAGGT	251502
Qy	96	PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe	115
Db	251503	TTTTTATCGTCAGGACGGTGACAAACAATTTATACGAGAGCGGTGCTTTTAACTC	251562
Qy	116	ThrGlnAspGly----PheLeuAsnAspProSerGlyPheThrLeuMetGly-----	Ser 132
Db	251563	GATAAAACGGTTACTACGTAACCTGCTAATAGTCTTAAGGTTCAAGGTGGAACCTCT	251622
Qy	133	ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe	152
Db	251623	AGCTGGATGATAAGGAATAAATACATCAATTCGGCAGCTTCGATTGAA---GATATT	251679
Qy	153	AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal	172
Db	251680	ATAATTCAGTATATTTCTAAGAACCTGCAAGGCTACTTCTCAGATAGATTTTAAATCC	251739
Qy	173	AsnLeuGlyAspSerThr-----AspLysThrGlnSerGluAlaAsnPro	187
Db	251740	AATTTGAATCTTCTGTCGGCGGTTCCGCTGACCAACCCCAAGAGAATACACT---	251796
Qy	188	TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsn	207
Db	251797	-----CGCATGATTAACGAT-----CCAGATCTCAAGATGAGA	251829
Qy	208	TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr	227
Db	251830	AGGGACATGTTACACGATTAATACATTTGACATCAAGGATCAAGAGATTTAA	251889
Qy	228	ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla	247
Db	251890	ATGGAATTTACAGGTTCTGATAACACTTGGAAAGCTAGATTGATGATGACGGATTCA	251949
Qy	248	MetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla-----	263
Db	251950	ACTCAACTTTCCGTTGATTTCTTGGAAACCGCGGCAAAACACTCAACTTCCAGGTAAC	252009
Qy	264	-----GlyLeuLeuMetSerGlyThrMetThrPheSerSer-----	275
Db	252010	ACTGAATTAGTAATTTGGTTTACTCCGATGGAAACTAGTATACGTTCTGACGGGTG	252069
Qy	276	-----AsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer---	290
Db	252070	GATTCGATGAACACGGGTAGTTTGAACGCAAAAGTTCTTTTCGGATTCCTCGGAATCCT	252129
Qy	291	-----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuPro	308
Db	252130	CGGATTCMAAATTTGATCTTAATCTTAGGAA---GCAGGAATGGTAAACGGAATCACT	252186
Qy	309	GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIle	328
Db	252187	CAATTCCTCTCTGACTTT-----	252204
Qy	329	LysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp	348
Db	252204	-----	252204
Qy	349	IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg	368
Db	252204	-----	252204
Qy	369	AsnGlySerSerSerThrArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuVal	388


```

/ gene="fliM"
3865 4899
/ gene="fliM"
/ note="flagellar switch protein"
/ codon_start=1
/ transl_table=11
/ protein_id="AAB61256.1"
/ db_xref="GI:2196898"
/ translation="MTSLVSDQIDQLLTALSSGDSASIEDARPISTRKITLYDPRP
DKPSKQMRILSLMHEPALTTSLSAQIRSMVHVVASVDQLTVEIRSIPTPTSP
LAVIMPLKGNVLEVPDPTISIDRLFGGQAKVORDLTDIENSVMGIVIRI
LANVRESQVQIDRLPELQGIETNPQAFQVPPSEMVLVLETKVGESEGMNFCIP
YITTEPISLKSQFSPSVRSSTQYMGVLRKLSLTDMDVVAEVSRLRSVRDIL
GLRVGDIIRLHDTHVGDPPFVLSIGNRKKFLCPQGVVGGKIAAQILERIBSTSEDPEE
LSADEEELYE"
4880 5230
/ gene="fliY"
4880 4883
/ gene="fliY"
4892 .>5230
/ gene="fliY"
/ codon_start=1
/ transl_table=11
/ product="fliY"
/ protein_id="AAB61257.1"
/ db_xref="GI:2196899"
/ translation="MSDGSISQABIDALLSGVMGADSPFAGASFGGLSQAQSEL
LQSPCANVGSFQPSLESIGTKTVSISPPREVISKEEFSRVPDMLNATVIDFGEF
AGDHLFLWAPF"
ORIGIN

Alignment Scores:
Pred. No.: 4,25e-25 Length: 5230
Score: 534.50 Matches: 154
Percent Similarity: 42.70% Conservatives: 80
Best Local Similarity: 28.10% Mismatches: 183
Query Match: 20.67% Indels: 131
DB: 1 Gaps: 14

US-10-009-823A-1 (1-502) x TPU28219 (1-5230)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
22 ATGATGGGTCACCTTTTTCAGGTGCTCTGATGAGAAATCATCAAGCCGCGCATGGAT 81
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 GTCAATGGGAACACGTCGGGACGTTTAACTACCGGTTTAAAGCGTGGCGGTGTTAAT 141
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 TTTCAGATCTTATTCTTCAGCACTGAGTGGCGGTGGCGTCCGATGAGAGTTGA 201
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 GGAGTAATCCCAAGGAAGTGGGATGGCGTGTGATTCAGACATCGATATCTGTTTAC 261
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
262 ACGCAAGTGCATCGCAACACGCGGTATCAATACGATGTGTCTATTACGGGAGTGT 321
QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
322 TTTTGTGTGAAAGTGGGAAAGACGCTTTTTCACCGCGCAGTGCCTTGGGTT 381
QY 116 ThrClnAspGlyPheLeuAsnAspProSer-----Gly 126
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 GATAATCGCGGACTCTCTGGAACCTTCGGAATGATGCGGCTTCAAGTTGGATGGCG 441
QY 127 PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu 146
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 CAGGACGTGGCGGAGCGGTTTAAATTAATTCCTCTGCACAGACGAGGATCTCGTTATC 501

```

```
Db      1327 ATTACCAAAAGGGTTTCAGCGGCGCAAGACGATTTCAGACATCAGACACCATGTTG 1386
QY      495 GlnLysAlaLeuGlutLeuLysArg 502
Db      1387 GATACCGTGTGATGTTTGAGCGC 1410

RESULT 15
AE001244/c
LOCUS   Treponema pallidum subsp. pallidum str. Nichols BCT 10-OCT-2003
DEFINITION the complete genome.
ACCESSION AE001244 AE000520
VERSION   AE001244.1 GI:3323005
KEYWORDS
SOURCE    Treponema pallidum subsp. pallidum str. Nichols
ORGANISM  Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
REFERENCE 1 (bases 1 to 26536)
AUTHORS   Fraser,C.M., Norris,S.J., Weinstein,G.M., White,O., Sutton,G.G.,
           Dodson,R., Gwinn,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
           Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
           Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
           Utterback,T., McDonald,L., Artiach,P., Bowman,C., Cotton,M.D.,
           Fujii,C., Garland,S., Hatch,B., Hurst,K., Roberts,K., Sandusky,M.,
           Weidman,J., Smith,H.O. and Venter,J.C.
           Complete genome sequence of Treponema pallidum, the syphilis
           Spirochete
JOURNAL   Science 281 (5375), 375-388 (1998)
MEDLINE   98332770
PUBMED    9665876
REFERENCE 2 (bases 1 to 26536)
AUTHORS   Fraser,C.M., Norris,S.J., Weinstein,G.M., White,O., Sutton,G.G.,
           Dodson,R., Gwinn,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
           Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
           Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
           Utterback,T., McDonald,L., Artiach,P., Bowman,C., Cotton,M.D.,
           Fujii,C., Garland,S., Hatch,B., Horst,K., Roberts,K., Watthey,L.,
           Weidman,J., Smith,H.O. and Venter,J.C.
           Direct Submission
JOURNAL   Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
           Medical Center Dr, Rockville, MD 20850, USA
FEATURES
     source
       1. 26536
           Location/Qualifiers
             /organism="Treponema pallidum subsp. pallidum str.
             Nichols"
             /mol_type="genomic DNA"
             /strain="Nichols"
             /sub_species="pallidum"
             /db_xref="taxon:243276"
             complement(100..303)
             /gene="TP0701"
             complement(100..303)
             /gene="TP0701"
             /note="similar to GB:L10328 SP:P08374 GB:M15266 GB:M24503
             PID:146978 percent identity: 27.59; identified by sequence
             similarity; putative"
             /codon_start=1
             /transl_table=11
             /product="DNA-directed RNA polymerase, putative"
             /protein_id="AAC65570.1"
             /db_xref="GI:3323006"
             /translation="MIFPMQQLIEFGNIYETCAATRAAFQIAAFCVPLDELGGKV
             VSAAAQQVFGSGVDYRIEPELQ"
             /gene="trRNA-Pro-3"
             /note="trRNA-Pro-3"
             /product="trRNA-Pro"
             /note="codon recognized: CCC"
             complement(473..1009)
             /gene="TP0702"
             complement(473..1009)
             /gene="TP0703"
             complement(1704..3824)
             /gene="TP0704"
             complement(1704..3824)
             /note="similar to GB:AE000783 percent identity: 40.76;
             identified by sequence similarity; putative"
             /codon_start=1
             /transl_table=11
             /product="single-stranded-DNA-specific exonuclease (recJ)"
             /protein_id="AAC65672.1"
             /db_xref="GI:3323008"
             /translation="MKERAKALDEGRVRELTARYSCSALEAILVRKVVADALLF
             HFERDLALPFPFRFHAMSAVDRLIAHERKEKVLIFGDRDADGI SATTLFLFALCA
             FGLTVCVRVPADPGLSCNADVEHAAAGTLITVDCGISNRAETAYARHRIQIDVL
             VIDHAAPELTPALIIINAKVPGAGYPHEFLSGCATAKVVTALRFTSGVGRPVVL
             LHVPTVGATIEAVKCNMEVDRLESTVPRMLPSHTRLGALFSGSVFVWDEOV
             VRTQAKVGFGEFWRFTDVRPMAAFPPQERASLLRMKDFSRIGARSKPMEEIS
             FINLFTFVKDTQLYGADELQVAISLSDMLVLAGENIRIVRIGLOMNPAPRV
             GLRELFALQIMGKKLQTVGVGWSIVPLINATGRGCPHRAVELFLMEASQALQAR
             KMVOLNEQRKLGSRARALVEPLRASLETYSNRLAVVCSDKTHRGVTGILANLSEA
             LRVPCLVICIMADGHVGLRSARGVHLFSLDPLADLFSDYGGHFAAGFSPISERI
             POLHRLMELVATIEFADESAQCEGFEDELATQMTRGLLTLVDRPEPYGEGNPPLC
             FLAKRLKIFASLFGRTERVHVHKLTLDAQVHKPFIYWGAGEKLAQFAVGVDVDAVF
             QVTRTFQGTCTPOLVIQDLWRSE"
             /gene="TP0705"
             complement(3967..6621)
             /gene="TP0705"
             /note="similar to GB:AE000783 percent identity: 40.76;
             identified by sequence similarity; putative"
             /codon_start=1
             /transl_table=11
             /product="penicillin-binding protein (pbp-2)"
             /protein_id="AAC65673.1"
             /db_xref="GI:3323009"
             /translation="MRSVCVLYLVLLCGLVGAALFGFLASTENIKQSERFTFRNP
             ALPRLIDIRGLDITESSDEKREIVSFADLSHLVHALLTREDRSFVTHGYSIKAI
             FRALVGLTNGALGGGTLTQQIAGLLYSRDSRSLRRKIKLMLWALHMERRYSKNEI
             MELYNRVTGALGGTGVGAAAFYFGHSVRQMSAAEALILKLSNPAHNPFPYPNR
             AQDRQVYLMETRLGVLSEQDESYHWAHFDYTRTTSSAFYABADKARFSEYV
             RROLDMMYGMTNLYQDGYVHTTCDLRHQLVAAQVEQVEQLAQHVKQSKLTHVAR
             RELYSNLSLGLVENVPLQHVGRMKAKSGAYRSTLGPLVNMALMFGDLNKLTL
             SDKASLAVRDEIARKEVGALENDTGYTLVGGSRFGASNOVIRATGILLQPGS
             VFKLPLVSAALDSKCLTNATQLHDPAQFNRNGVSIYPNYGGKQGVGLAWKLAOS
             LNIPIAIRVLDVMVGFDAVIQRAATLLHITDRQEIERTFPFVPLALGVVALRPIQLARA
```

PAFCNGCKAVEPIAVSRVEDRLGRVILDPEREVRARLRQAQGAATOLISAENALMTN
 MLEKTVMTGLAVASERGRAFTYODPATGRSFVMPVAGKTGTTTONSDAWAGYSYY
 TAVLWFGDGLSGHSTGATLAGEPWAFMRATHQDHPRAFRVPRPETGLMYQAVC
 AKSLGHTPYCGDELVSFLPGTVPFRFSGHESGKMLBEAKERLKSQSGYSTGFWP
 AVSEGLHIDPRVPEPDPGSPGVQGEVLVPSNPFAPVHPQEAFLDAPHEVPVPSASD
 PSRDAPESESDVQDDPARGVQSLDTGP"
 complement (6918..7844)
 /gene="TP0706"
 complement (6918..7844)
 /gene="TP0706"
 /notes="similar to GB:AE000783 percent identity: 36.30;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC65674.1"
 /db_xref="GI:3323010"
 /translation="MVRAAVCSCLVCVHPAVCGPYPLIAQLQGQDALFRQYSDVVR
 AARVALAQGRGNDLPRFYAVYKXADTIIRIAARCGIPYDAVASINRIETLHLE
 GRTLLPTVGLVSDADPHLPFLERLIVALKKGQSPFLSPGTAQTTHTRKRVV
 CAPALPDGVTRAFELPKPYRFPPLASRLTSGFGARKSPETGRLSVHPCGIDLAAPGA
 LVACASGQVATYRNLKYLQHTDGRHSLYGHLASAVRVVQOKLSVGAVIGNV
 GSTAGTGPLHFEVRAGVQNPFRFMEXR"
 complement (7906..8385)
 /gene="TP0707"
 complement (7906..8385)
 /gene="TP0707"
 /notes="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0707"
 /protein_id="AAC65694.1"
 /db_xref="GI:3323030"
 /translation="MRGTAYAVSVGVSACTCTOVAVQLASWFGMRKRCVEVR
 RRCALMAYVRVIRIGLAMLAVLSFALGCALVFLMALAVHRFRVSVLSGLLYGG
 GAVLGLRRVCNALSARVRRARRAAQEPCLVQAGEVGANGUSSAEVRPAQEC"
 8451..8966
 /gene="TP0708"
 8451..8966
 /gene="TP0708"
 /notes="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0708"
 /protein_id="AAC65695.1"
 /db_xref="GI:3323031"
 /translation="MDVQSRFSCAAVGLKVPATAAGAAFFLSIATAAARNARVY
 VSVARATVLAGCATAYARALLAVVFDLLVQEDGKMPIDPHVDLTLDVVEPFAS
 PGDVQVETDSDSFDLSIPTGELGELGYSFSPSPSFEDKTSIDLVDGLVDLDELTKRMA
 RAIRTVLSQDT"
 9022..9810
 /gene="TP0709"
 9022..9810
 /gene="TP0709"
 /notes="similar to GB:J03169 SP:PI7211 PID:153486
 PID:153487 percent identity: 47.79; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase sigma-28 factor"
 /protein_id="AAC65675.1"

Alignment Scores:

Pred. No.:	2,91e-24	Length:	26536
Score:	534.50	Matches:	154
Percent Similarity:	42.70%	Conservative:	80
Best Local Similarity:	28.10%	Mismatches:	183
Query Match:	20.67%	Indels:	131
DB:	1	Gaps:	14

US-10-009-823A-1 (1-502) x AE001244 (1-26536)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetIysThrHisSerThrGlyLeuGly	20
Db	26471	ATGATGCGCGTCACTTTTTCAGGTGTGTGTATGATGAGTATCAACACCGCGAT	26412
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal	40
Db	26411	GTCATGTGGAAACAGTCGCGAAGCTTACACTACCGGTTTAAAGCGTGGCGGTAA	26352
QY	41	PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer	55
Db	26351	TTTCAAGATCTTATTTCTCAGCAACTGAGTCGCGTCCGCAATGAAGAAGTTGGA	26292
QY	56	SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe	75
Db	26291	GGAGTGAATCCCAAGGAAGTGGGATGGCGTGTGATTCAGCATCATGACTGTTCAC	26232
QY	76	ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly	95
Db	26231	ACGCAAGGTGCTGCAACACGCGGTATCAATACGATGTGTCTATTACAGGAGTGT	26172
QY	96	PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe	115
Db	26171	TTTTTTGTGCTGAAAAGTGGGAAAGACGTTTTCACCGCGCAGGTGCTTTGGGGTT	26112
QY	116	ThrGlnAspGlyPheLeuAsnAspProSer	126
Db	26111	GATAATGCGGCACCTCTCGTGAACCTCGAATGGTATGCGGTTCAAGTGGATGGCG	26052
QY	127	PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGlu	146
Db	26051	CAGCAGCTGGCGGGAGCGTTTAAATTAATTCCTGCACAGACGAGGATCTGTTATC	25992
QY	147	ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr	166
Db	25991	CCCAATTGGG-----CAAAGATAGTAGTCGCGAGCAGCAGCAGCAGCAGCAGC	25956
QY	167	AlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn	186
Db	25955	ACTGTTCACTATGCTGTATTAATTAGACAAGCGTCTGCTGAGCTTGTGCAGATGCGAAC	25896
QY	187	ProTyrPheAlaLeuLeuGluSerThrLysGlyAsnGlyThrProIleSerThrSer	206
Db	25895	GAAGCGACGTGCTAAGTCCACGTG-----ACAACT	25863
QY	207	AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle	226
Db	25862	GACTTT-----CAAGTGTATGATAGCTTCGGGACGACGATACGTTG	25821
QY	227	ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal	246
Db	25820	CAGATTAACTTTTCGCTGTCGCGGGACG-----AACAACTCAGTGGCAGCCATGTC	25767
QY	247	AlaMetAsnPro	253
Db	25766	GCACTGGATCCGGGACAGAGGTAGATACGAAACGGTGTAGGGTGGGACATCTGAC	25707
QY	254	GlySerAla	260
Db	25706	GGTGGGCAACACCTTTATTGATAAATTTTGACACCTTCGCTGCTGAGTACT	25647
QY	261	AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys	280
Db	25646	GACTGACGG-----	25635
QY	281	AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaThrGlnPro	300
Db	25634	AACGTGACCGGT-----CCTACCGGACAGGTG-----	25608
QY	301	AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln	320
Db	25607	-----CTCCTTGAAGCGTGTACGATGTTGCGTGGC-----	25575

Qy	321	ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrrAlaGlyAlaProAla	340
Db	25574	-----AATCCGGACATCGAGCGAGGTT	25551
Qy	341	SerAlaAlaAlaIleGlyThrAspIleGlyLeuProSerMetMetProIleGlnThr	360
Db	25550	ACCGCCACCGCTTTCACGCTCACTTGGGTGAATT-----	25515
Qy	361	SerSerGlyAsnSerThrAlaArgAsn-----GlySerSerSerThr	374
Db	25514	-----GGCACCGCGCCCAATACGATTACCGATTGCTGAAGCGAGTACTAC	25467
Qy	375	ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer	394
Db	25466	AAAGCCTACCGCAGCAGCGTTACCGCATTTTGGAAATTTTAAATATAGATCAA	25407
Qy	395	GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro	414
Db	25406	AGCGGTGTCTACTCTGTGTATCAATGGGTGAGCGACACATTTGGCCAGCTCGCA	25347
Qy	415	LeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAla	434
Db	25346	CTTGCAGGATTTCAAAATCAAGTGTCTTGAGAGCGCAGGAGAGAACACCTACGTACAA	25287
Qy	435	ThrLeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu	454
Db	25286	TCGAACAACACTCAGCGATAGCGAATATTAGCACTCGCGGGGTGATGGGAGGAAAGTTG	25227
Qy	455	SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle	474
Db	25226	ATTGCAGGGACACTTGAGATGACCACTAGATTAAACCGATCAATTTACGGATATGATC	25167
Qy	475	IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu	494
Db	25166	ATTACCCAAAAGGGTTTACAGCGGGCGCAAGACGATCAGACATCAGACACCATGTTG	25107
Qy	495	GlnLysAlaLeuGluLeuLysArg	502
Db	25106	GATACGGTGTGTGATTGAAGCGC	25083
RESULT 16			
BX321857 313050 bp DNA linear BCT 23-APR-2003			
LOCUS Nitrosomonas europaea ATCC 19718, complete genome; segment 2/10.			
ACCESSION BX321857 AL954747			
VERSION BX321857.1 GI:30138174			
KEYWORDS complete genome.			
SOURCE Nitrosomonas europaea ATCC 19718			
ORGANISM Nitrosomonas europaea ATCC 19718			
REFERENCE 1 Chain, P., Lamerdin, J., Larimer, F., Regala, W., Land, M., Hauser, L., Hooper, A., Klotz, M., Norton, J., Sayavedra-Soto, L., Arciero, D., Hommes, N., Whittaker, M. and Arp, D. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph Nitrosomonas europaea J. Bacteriol. 185 (9), 2759-2773 (2003)			
25586410			
12700255			
REFERENCE 2 (bases 1 to 313050)			
AUTHORS Larimer, F.			
TITLE Direct Submission			
JOURNAL Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas genome consortium, the DOE Joint Genome Institute, Production Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA, and the Genome Analysis Group, Oak Ridge National Laboratory, 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA; larimerf@ornl.gov			
REMARK Nitrosomonas genome consortium			
FEATURES Location/Qualifiers			
1..313050 /organism="Nitrosomonas europaea ATCC 19718"			
misc_feature			
/mol_type="genomic DNA"			
/strain="ATCC 19718"			
/db_xref="taxon:228410"			
complement (132..503)			
/locus_tag="NE0270"			
complement (132..503)			
/locus_tag="NE0270"			
/codon_start=1			
/evidence=not experimental			
/transl_table=11			
/product="hypothetical protein"			
/protein_id="CAD84181.1"			
/db_xref="GI:30138175"			
/translation="MPEQITRKASPFCLRTTPERTLLEREAAGLPGEVIHQVDFE NVRRGRNQPVKDHLLSQLLGELGRSLANNLNQAPAAACGLNLATPEVKTSL NACADIRHRETLMKSLGNR"			
complement (885..1253)			
/locus_tag="NE0271"			
complement (885..1253)			
/locus_tag="NE0271"			
/function="InterPro IPR002559"			
/codon_start=1			
/evidence=not experimental			
/transl_table=11			
/product="Transposase IS4 family"			
/protein_id="CAD84182.1"			
/db_xref="GI:30138176"			
/translation="MFSLSFCQAGADPEGRKLLKTLNKGFSDFTHVIMDKAYEGDETR QLVLDGMIPVPPKANRVSFWEYDVEYKRNVEVERLFRRLKGFRRIFSRFDKDSV FRFIHFALIVDHLISVNRP"			
complement (887..1698)			
/note="ISNE5"			
complement (1253..1660)			
/locus_tag="NE0272"			
complement (1253..1660)			
/locus_tag="NE0272"			
/function="COGs COG3293"			
/codon_start=1			
/evidence=not experimental			
/transl_table=11			
/product="possible transposase"			
/protein_id="CAD84183.1"			
/db_xref="GI:30138177"			
/translation="MEISASQFKLIENLLPIQRGNVTLNSLEVLNAILYVAEHGCKWR GLPKVGNWHSIYTRANRWARGVLDVFLALQONKLIQLEVDHMSLDSTIVKVPDGTGALKKTVFKLSVNEHGAGLPKFTWSRQMPER"			
complement (1853..2488)			
/gene="cyc"			
/locus_tag="NE0273"			
complement (1853..2488)			
/gene="cyc"			
/locus_tag="NE0273"			
/function="InterPro IPR000345:IPR002329:IPR003088 COGs COG2863"			
/codon_start=1			
/evidence=not experimental			
/transl_table=11			
/product="Cytochrome c, class IC: Cytochrome c, class I"			
/protein_id="CAD84184.1"			
/db_xref="GI:30138178"			
/translation="WKIIGTVAIAIALTFSGFVFAESPAAGDVKEKGEIAGICAGCH NADGNSAIPYILAGYPGYIAKQLNDFKVEGETVKRDNQIMAPVAILSQDDIEN LAAYSQKQPGFASIVETGKLYQGNLENSIPACSSCHNGQGPPIPPRI DQHPAYTLSQLQAFROGTRKNDTNTMTQIVSRMSQEMQAVSEYIATLR"			
complement (2423..2488)			
/gene="cyc"			
/locus_tag="NE0273"			
/note="Signal predicted by SignalP 2.0 HMM (Signal peptide probability 1.000) with cleavage site probability 0.999 at residue 22"			
2595..3197			
/locus_tag="NE0274"			
2595..3197			

QY 204 SerThrSerAsn-----TyrSerTyAlaGlnProMetArgValTyAspGlnGln 220
 Db 31592 AATGTAGTAAATCCAGATACCTATACCCATTTACTTCGGGAGAAATTTTGCACAGTTT 31651
 QY 221 GlyAsnSerHisAspIleThrValTyPheAspGlyAlaProSerSerThr---GlySer 239
 Db 31652 GGTAAATCCCATATATTAATTTATTTCCAGAAAACAGCTGCATCCACTTGGGGCGTA 31711
 QY 240 LysThrPheGluTyLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly 259
 Db 31712 TATACAGCCGCGATGGTACGTAGTAGTACCTCT-----ACTGGATTATCTGCT 31762
 QY 260 ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeu 279
 Db 31763 ACGGATAGC-----CCAACTACTTTCAGCTTAATCCATCCCGG----- 31801
 QY 280 LysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGln 299
 Db 31802 -----ATCATGATCACCACCGGTCAGATGACATTAATGTG----- 31837
 QY 300 ProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIle 319
 Db 31838 -----GCTGCTTCGTACTGCTGCTGCGGCGGTT 31864
 QY 320 GlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaPro 339
 Db 31865 TCACCGCTGCTTTTCAGCTTGATTTAAC----- 31894
 QY 340 AlaSerAlaAlaLalleGlyThrAspIleGlyLysLeuProSerMetMetProIleGln 359
 Db 31894 ----- 31894
 QY 360 ThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTySerGln 379
 Db 31895 -----CGTTCACCCAGTTTGCAGCAAAATTCGGGTGAATGATGATGAGCCAG 31942
 QY 380 AspGlyTyProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGln 399
 Db 31943 GATGATTTGCTTCGGGTGCGCTGCGGTCAGACATTTCTGATGATGCGGATCAGCA 32002
 QY 400 GlyLysTySerAsnSerGlnValValAspPheTyAsnIleProLeuAlaArgPheThr 419
 Db 32003 GCAGTTACTCCAAATACGTCACGCGCAGTCAGTGGTCACTGTGTCAGGTTTTC 32062
 QY 420 SerGluAspGlyLeuArgArgGluGlyAsnAsnHisTySerAlaThrLeuAspSerGly 439
 Db 32063 AATCTTAACGCTCTCAAGCGCTGCGGAAGCAATCAGTGGGTGAGACGCGTGAATCCGA 32122
 QY 440 GlyProGluPheGlyLeuProGlyThrSerAsnTyTyGlyLysLeuSerValAsnGlnLeu 459
 Db 32123 TTGCGGTGTGGGGACTCGGAGACGGGTGCTGCTGCTCACTCAATCTTCAGCAGTG 32182
 QY 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGly 479
 Db 32183 GAAGAATCCAAATGTTGATCTGACGACGCGAGCTGGTCACTGATCATGACGCGCGGTC 32242
 QY 480 PheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGlu 499
 Db 32243 TATCAGGCCAATGCCAGACCATCGAGACAGATGCGATGATGATGACACCATCATGAAT 32302
 QY 500 Leu 500
 Db 32303 ATC 32305
 RESULT 17
 AEO13102/c 10303 bp DNA linear BCT 10-JUN-2004
 LOCUS AEO13102
 DEFINITION Thermoanaerobacter tengcongensis MB4, section 129 of 244 of the complete genome.
 ACCESSION AEO13102 AEO08691
 VERSION AEO13102.1 GI:20516444
 KEYWORDS

SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 gene
 CDS
 misc_feature
 RBS
 gene
 CDS

Thermoanaerobacter tengcongensis MB4
 Thermoanaerobacter tengcongensis MB4
 Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 Thermoanaerobacteriaceae; Thermoanaerobacter.
 1 (bases 1 to 10303)
 Bao,Q., Tian,Y., Li,W., Xu,Z., Xuan,Z., Hu,S., Dong,W., Yang,J.,
 Chen,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y., Ling,L.,
 Tan,H., Chen,R., Wang,J., Yu,J. and Yang,H.
 A Complete Sequence of the T. tengcongensis Genome
 Genome Res. 12 (5), 689-700 (2002)
 21992816
 11997336
 2 (bases 1 to 10303)
 Bao,Q., Xu,Z., Hu,S., Dong,W., Chen,Y., Wang,J., Yu,J. and Yang,H.
 Direct Submission
 Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and
 Bioinformatics Center, Institute of Genetics and Development,
 Chinese Academy of Sciences, Beijing Airport Industrial Zone B6,
 Beijing 101300, China
 3 (bases 1 to 10303)
 Li,W., Xuan,Z., Yang,J., Ling,L. and Chen,R.
 Direct Submission
 Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of
 Sciences, Beijing 100101, China
 4 (bases 1 to 10303)
 Tian,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y. and Tan,H.
 Direct Submission
 Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy
 of Sciences, Beijing 100080, China
 Location/Qualifiers
 1. 10303
 /organism="Thermoanaerobacter tengcongensis MB4"
 /mol_type="Genomic DNA"
 /strain="MB4"
 /db_xref="taxon:273068"
 complement (24..1284)
 /gene="FlgE3"
 /locus_tag="TTE1433"
 complement (24..1271)
 /gene="FlgE3"
 /locus_tag="TTE1433"
 /note="Best Blast hit = gi|2494528|sp|Q474767|FLGE_BORBU
 FLAGELLAR HOOK PROTEIN FLGE gi|7463172|pir|C70135
 flagellar hook protein (flge) homolog - Lyme disease
 spirochete gi|1165267|gb|AA85606.1| (U43739) FlgE
 [Borrelia burgdorferi] gi|2688183|gb|AAC66665.1|
 [Borrelia burgdorferi] flagellar hook protein (flge) [Borrelia
 burgdorferi], score 278, E-value 1.00E-73"
 /codon_start=1
 /transl_table=11
 /product="Flagellar basal body and hook proteins"
 /protein_id="RA024655.1"
 /db_xref="GI:20516445"
 /translation="MLRSMYSALSGLOAHOARLDVIGNNTAVNTVGYKSRMTFKEI
 FSOTIKGAPQGGGTNPQOIGLVALASIDTLFTRGGAQRTDNDIDLSIDNGFF
 IVSGESILYTRAGNFDNGDLVTPDGKVLGMMSTDGKVTNTDNLVPSLKNW
 FGMEPSYTTQIEIGGNLNASTGICQSIYNIIVYDTQGGSHVATITTRQDPNTWNS
 VSSDDPTSSVSVAGSTAITFGADKIAPIATQATGTLTNNMTAVTNAQIPVNIIDS
 KLTWSEITDLRELKNGNEAGSLQSIDKYGVSIGVNGRQVIGQIALADFQNP
 MGLKYNHTFINTVNSGEPMGAGSGSTRGINSFPLENSVNLANFEVDMITTQRG
 FQANAKVITVSDEMLQDLVNNKR"
 complement (1167..1259)
 /gene="FlgE3"
 /locus_tag="TTE1433"
 /note="Pfam match to entry flg_bb rod, Flagella basal body
 rod protein, score 58.3, E-value 1.30E-13"
 complement (1280..1284)
 /gene="FlgE3"
 /locus_tag="TTE1433"
 complement (1352..1712)
 /locus_tag="TTE1434"
 complement (1352..1702)
 /locus_tag="TTE1434"

```
/note="Best Blastp hit = gi|10175070|dbj|BAB06169.1|
(AP001515) BH2450
unknown [Bacillus halodurans], score 75.5, E-value
2.00E-13"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM24656.1"
/db_xref="GI:20516446"
/translation="MKVHFLNSGPSIQKPSYGEAFKQVFEKLAIEKFSKSLLR
MEANVKITDPEYKYLLEAVEKAKSSGIKSLIIMDKAFVNLKSRVTITAMDESM
KNSVFTNDGAVII"
RBS complement(1708. .1712)
gene complement(1712. .2122)
CDS /gene="FlgD"
complement(1712. .2107)
/gene="FlgD"
locus tag="TTE1435"
locus tag="TTE1435"
/note="Best Blastp hit = gi|7459749|pir|D70135 flagellar
hook assembly protein (flgD) homolog - Lyme disease
spirochete gi|1165268|gb|AAH85607.1| (U43739) YLXG
[Borrelia burgdorferi] gi|1196325|gb|AAH51417.1| (L76303)
flagellar hook assembly protein [Borrelia burgdorferi]
gi|2688182|gb|AAC66664.1| (AE001137) flagellar hook
assembly protein (flgD) [Borrelia burgdorferi], score
85.1, E-value 3.00E-16"
/codon_start=1
/transl_table=11
/product="Flagellar hook capping protein"
/protein_id="AAM24657.1"
/db_xref="GI:20516447"
/translation="MNVNTYSLVPSNKNVTKSKLGDPLMLLVTLQKNQDPLN
PVDDKEFLAQFSALEQNLNENFNSKALIGKNVYATINDGNGYKEVFGKV
DAVRENGQFLKVGKIDIPDAVRAVTE"
RBS complement(2118. .2122)
gene /gene="FlgD"
locus tag="TTE1435"
complement(2125. .3349)
locus tag="TTE1436"
CDS complement(2125. .3336)
locus tag="TTE1436"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM24658.1"
/db_xref="GI:20516448"
/translation="MSEQVVQSLLPFLKAVDSKMAKGKNGKVSFLDILLSELTE
TRENKENPLNKKEFLSKQKGNVEIFQEKREDQIMEDIANFIPALMQITLL
SLETISSEKGFKEKREKLEVALQGFIERNFTPEIAKIDSLEKFNIEISPE
VIERHILKARVDLPDLNOKDFAEENQVQNDQKIDKEAPFAAKAEKBE
KTEKSPDVQKQEVFFNEQKPVSNLTYSIKKSNPDVRLFRQIVDVFVAKEGAS
SVTVNLKPEILGLQISLSDIGNIVATIVTESEKTKHOIESNLSLQALDLKGI
ESNVAVDKNFQFTPQYTGSGNYQSGYGCENENLRVFIHVGSKYLPSELDITQNY
YFKDGHLDLMA"
RBS complement(3345. .3349)
gene /locus tag="TTE1436"
complement(3352. .3944)
locus tag="TTE1437"
CDS complement(3352. .3930)
locus tag="TTE1437"
/note="Best Blastp hit = gi|10175073|dbj|BAB06172.1|
(AP001515) BH2453
unknown conserved protein in B. subtilis [Bacillus
halodurans], score 62.8, E-value 3.00E-09"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM24659.1"
/db_xref="GI:20516449"
/translation="MQNEQVQEKRRKAKWIIALILVILSSVLVYNIGSQVFL
TKQISKIPLKGVSPKTEDQDAVIAKLNEELKGEATISYEKIEQNAQIBELX
```

```
KQIEELQNEKUTLQKQFESRONSUKDIATYYQNMDPQNAQAQLNMSDEVIKILTSM
NKDASKIMEALKPKAKAKITNMLLKATWVP"
complement(3940. .3944)
locus tag="TTE1437"
gene complement(3943. .4397)
CDS /locus tag="TTE1438"
complement(3943. .4386)
locus tag="TTE1438"
/note="Best Blastp hit = gi|7462657|pir|F72368
hypothetical protein - Thermotoga maritima (strain MS88)
gi|4981024|gb|AAD35597.1|AE001727_9 (AE001727)
hypothetical protein [Thermotoga maritima], score 57,
E-value 1.00E-07"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM24660.1"
/db_xref="GI:20516450"
/translation="MKKFEFTLOSVLNLEKQSEKIEKENLAKIMKEIEREREKLENLK
KHEQVTKRAKEVEEGTLMYKLAETAYIMKIREMIEKOANYTLKLEKAEKIREGL
LKVSKEKKALENLKERQFSEYLYLLNLEQSRVIDEHVSYKAKSY"
gene complement(4392. .5716)
CDS /gene="Flil"
complement(4392. .5705)
locus tag="TTE1439"
locus tag="TTE1439"
/note="Best Blastp hit = gi|10175075|dbj|BAB06174.1|
(AP001515) flagellar-specific ATP synthase [Bacillus
halodurans], score 529, E-value 1.00E-149"
/codon_start=1
/transl_table=11
/product="Flagellar biosynthesis/type III secretory
pathway Anpase"
/protein_id="AAM24661.1"
/db_xref="GI:20516451"
/translation="MSSGVLEKYSKILREKRLVGYGKVSQVIGLTIESVGPLSNIGE
ICYIKTDGNEVLAEVYKVEKLYMLPLNGMEIGPGKVIATGOTLKVNVGKSLIG
RVLDLGNPIDGKPLKYKSIIPVNTPPDFLEKRIREVPLGKIAIDGLLLTCGKQ
RIGIFAGSGVCKSTLLGMARNADNLVIALIGERRENEFTEKOLBEGELKRSYV
VWATSDPALVRVKGAVTATAIARYFDQGLDVLMMDSITRFAMAQREIGLSIGAP
VSRGYTSPVSFVLPKLLERSGSSQKSIATLYLVNDGDLNPSIADAVRGILDGTV
LSRLKANRHPAIDVLVASVRSVINDIITDEHEKELVAKFKILATYAEADLINIGY
NFGNPKIDAEIENKERNKSFRLKROKIDEAYDFETIKKLAKLSUER"
RBS complement(4393. .4397)
misc_feature complement(4623. .5432)
/gene="Flil"
Alignment Scores:
Pred. No.: 2,33e-24 Length: 10303
Score: 528.50 Matches: 148
Percent Similarity: 43.25% Conservative: 77
Best Local Similarity: 28.91% Mismatches: 180
Query Match: 20.44% Indels: 107
DB: Gaps: 12
US-10-009-823A-1 (1-502) x AE013102 (1-10303)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 1271 ATGTTAAGTCATGTTATTCGCCATTTCAGGTTTACAGGCTCATCAGCAAGGTTGGAT 1212
QY 21 ThrValSerAsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValVal 40
Db 1211 GTTATTGAAATAATAGCAATGTAATACAGTGGGTATAGCAAGCAGGATGACT 1152
QY 41 PheGlnAspLeuPheSerGlnAspLeu-----AlaIleGlySerThrGly 55
Db 1151 TTTAAGAGATATTTCAGCCAGACCATAAAGAGGACTTCGGCTCCTCAAGGAATGGCGGA 1092
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
```


demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).
 Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.
 Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

```

FEATURES             Location/Qualifiers
     source            1..25022
     mol_type="genomic DNA"
     strain="PAO1"
     db_xref="taxon:208964"
     complement(69..770)
     gene="brag"
     locus_tag="PA1070"
     complement(69..770)
     gene="brag"
     locus_tag="PA1070"
     note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane-associated; Subcellular localization confidence: Class 1"
     codon_start=1
     trans1_table=1
     product="branched-chain amino acid transport protein Brag"
     protein_id="AAG04459.1"
     db_xref="GI:9946988"
     translation="MISFDKVTSTYKIQALHDVSVVEKGEIVTILGANGACKSTLL
FTDQDDQVQASGIRVEGELVGLPSTMRKSIADVPEGRVFSRLTVEENLAMGGF
SLGLAPIIIQOIPEIIEQLRREGVTFLVQANQALKADRAYVLENGRIVMHDITGA
ALLNPKVRDAYLGG"
     complement(773..1540)
     gene="brag"
     locus_tag="PA1071"
     complement(773..1540)
     gene="brag"
     locus_tag="PA1071"
     note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane-associated; Subcellular localization confidence: Class 1"
     codon_start=1
     trans1_table=1
     product="branched-chain amino acid transport protein Brag"
     protein_id="AAG04460.1"
     db_xref="GI:9946987"
     translation="MSRPILEVSLTMRFGGLLAVNGVNLKVEEKQVSMIGPNAGK
TVFNCUTGYQPTGGHILRIDGEIEIQLPGHKIARKGVNRTFONVLFKEMTAVENILL
VAQHRHLNTFPLAGLFTPAFRSREAMEYAAHLEEVNLTFAFRSAGTLAYGQQR
RLEIARCMTRPRILMDPEAGLNPKETDLDLAKLAKLSEHNVTLLIEDHDKMLVM
SISDHIVVINGQAPLADGTPQIQRNDPDKVAYLGEA"
     complement(1537..2790)
     gene="brag"
     locus_tag="PA1072"
     complement(1537..2790)
     gene="brag"
     locus_tag="PA1072"
     note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 1"
     codon_start=1
     trans1_table=1
     product="branched-chain amino acid transport protein Brag"
     protein_id="AAG04461.1"
     db_xref="GI:9946989"
     translation="MSQSLRRALFSAALLVIIVSYPIGLKLRVTKLEVLGADAGTL

```

```

WTIAAALAMFVWQLPRDRIPLKIGRGVGVKNGSLKNFSLSPSTQRMVAVLALVVVA
FWPFPSRAGVADIATILIVMLGINTVWGLAGLLDLGYGVGFVAVGATYVALLAE
YAGGFWTALPACGMALFGLLGRPVLRGLRGLAYLNTGLGGEIIVRILLRNTEIT
GPGNGISIKPPTLFGLTERRAPEGMOTHEPFGIAYNTNYKVLIVVALLVLLA
GLVINLRMPFIRGAWREALREDEVACALGINFTIVKLSAFTIGASAGAGSFFAAR
QGLVTPSPFTIESAMILAIIVLGGMGSQLGVILAAVWVLLQEMRGFNEYRMLIFGL
TMIWMTWRPQGLLPQRPHELEKP"
     complement(2787..3710)
     gene="brag"
     locus_tag="PA1073"
     complement(2787..3710)
     gene="brag"
     locus_tag="PA1073"
     note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 1"
     codon_start=1
     trans1_table=1
     product="branched-chain amino acid transport protein Brag"
     protein_id="AAG04462.1"
     db_xref="GI:9946989"
     translation="MPEIYHVLOLVNGLTVGSTYALIAIGTVVYGIIGMINFAHGE
VYMGSYIATITLTAAMGLDSVPLMLAAFAASIIVTSAGFYSIERVAVRPLRGN
RLIPLISAIGMSIFLQNAVMSLQDSKEKAIPTLLPGNFVGESSMGVIVSYMQLIF
VITFLVPGULTFISRSRLGRACACAEEDUKNTLLGINSINLIATLTFVIGALAAVA
AVLLMGQVGINPGIFLAGIKAFKTAALVGGISIPGAMILGGLLGVAREAFGADVPGD
QYKQVAFGLLILVLLFRPTGILGRPEVEK"
     complement(3971..5092)
     gene="brag"
     locus_tag="PA1074"
     complement(3971..5092)
     gene="brag"
     locus_tag="PA1074"
     note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: Periplasmic; Subcellular localization confidence: Class 1"
     codon_start=1
     trans1_table=1
     product="branched-chain amino acid transport protein Brag"
     protein_id="AAG04463.1"
     db_xref="GI:9946990"
     translation="MKKGTORLSLFAAMAIAAGFASYSMAADTIKIALAGPVTGPVVAQ
IGDWKAGALMAIEQINKAGVNGAQLEGVYIDDACDPKQAVAVANKVNDGVKVFVG
HVCSSSTQPADTIYEDSGVLMITPSATPEITSRGYKLIPTRTIGLNNQGPVAGKFA
ERYKDKTIAVLHDKQOYEGEGATEVKTVEDAGIKVAVFGLNAGDKDFNALLSKLKK
AQGVQVYFGYHPMGILLKQAKOGLDARFMGPEGVNSEITAIAGDASEGMLATIP
RAFQDPQNKALIDAFKAKNQDPGSIQVLPAYSAVTVIKGIKAGADPEKVAEALR
ANTETPTGNLGEDEKDKLKNFDFTVYVHKHDA"TRTEVK"
     complement(5391..5708)
     gene="brag"
     locus_tag="PA1075"
     complement(5391..5708)
     gene="brag"
     locus_tag="PA1075"
     note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
     codon_start=1
     trans1_table=1
     product="hypothetical protein"
     protein_id="AAG04464.1"
     db_xref="GI:9946991"
     translation="WTEQSTLYAKLLGETAVISWAEIQPFARFAGALLQVDAALDLVE
VAEALAGDDREKVAAMLGGGLSKVGDAAKDFLERDPTLWAVVVPWVVIQERAEKA
TLH"
     complement(5775..6137)
     gene="brag"
     locus_tag="PA1076"
     complement(5775..6137)
     gene="brag"
     locus_tag="PA1076"
     note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"

```

Alignment Scores: 7,75e-24 Length: 25022
 Pred. No.: 527.50 Matches: 151
 Score: 45.84% Conservative: 86
 Percent Similarity: 29.21% Mismatches: 205
 Best Local Similarity: 20.40% Indels: 75
 Query Match: 16
 DB: 1

US-10-009-823a-1 (1-502) x AB004539 (1-25022)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 DB 8000 AGTTTCAACATCGGCTGAGCGGCATCCAGCGGCTCTAGCGGCTGAACGTCACCGGC 8059
 QY 24 AsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
 DB 8060 AACACATCGCCACCGCGCCGAGCTTCACGCAATCCCGCGGAGTTCGCCGAC 8119
 QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 DB 8120 GTCTACGCGCTCGGTG--CTGGGTTCG-----GGCAGCAACCGCGAGGCG 8164
 QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlnAlaPheGluProGly 83
 DB 8165 AGCGGGTGTCTCTCGGACGCTCGCAGATGTTCAACAGGGCAACATCGACTCGACC 8224
 QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPheGlnValThrLeuGluAsp 103
 DB 8225 AACAGCGTGTGACCTGCGCATCAACGCGCAACGGCTTCTGTCACCAACCAACGCGG 8284
 QY 104 LysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
 DB 8285 CGCATCAGCTACACCGCGCGGCTACTTCAATACCGCAACGAGGATTTCATGTCGAC 8344
 QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
 DB 8345 AACACGGCTACCGCTCGCGGCTATCGCTCGGCGCGCAACGCGCATGTCAGAACGCGC 8404
 QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
 DB 8405 GTGGTACCACGCTCAAGTCGAGCGCGCAATCAGCGCGCGCGACCCACCTCGAGCATC 8464
 QY 163 LysThrSerThrAlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGln 182
 DB 8465 CAGCAGTCGACACCTCACTCACTCGAGCTGAAG-----CCG 8500
 QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202
 DB 8501 CCGACCGTGCAGCGCTTC-----GATCGG 8524
 QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsn 222
 DB 8525 TCCGACCGCTACCTCACTCACTCGCTCTTCGCTGGGCTATGACAGCCAGGCGCAAC 8584
 QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
 DB 8585 TCCACACCATGACCGAGTCTTTCATCAAGACGCGCGACCGCAATGCGACCGCGCG 8644
 QY 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
 DB 8645 ATTCCGGAGAACGCTGGACCATGAAAGTGTGATGACGCGCTCAATCGCTCGATCCG 8704
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 8705 TCGAACAGACCGCG-----ATGAGCTTCAACGTCACCTTC 8740
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
 DB 8741 GACGCCAGCGCCAGATGACCTCGGTTCGGGCG-----CCGACGCGCAGCAGCGGG 8794
 QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304

DB 8795 CCGGGCTTCAGCATCGACGCGACCAACAGTGTATCAGTTCAGTTCGGGC----- 8845
 QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
 DB 8846 ACTGGCAATCGCGGACTCC-----GGCAGCGCTGGATTCCGGCGGCGCTCG 8893
 QY 325 AspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 344
 DB 8894 GAC---GGCAAGACCCCGCGACCTACGCTGGAAACGCGCGACCGGTGCCGACGGCG 8950
 QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
 DB 8951 ATCTCTTCGACATCGCCAG-----ACCACCCAGTAC 8983
 QY 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGln 384
 DB 8984 TCCACCGCTTCGCCCGCAGACCAACCGATC-----CAGGACGCTACACCAAC 9031
 QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
 DB 9032 GGTGAGTTCGACGCGCTGGAAATCGACACCGGGGTGATCTTCGCCCGCTACACCAAC 9091
 QY 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
 DB 9092 GCCAGTCCAAAGTCGAGGCGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9151
 QY 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
 DB 9152 AGCGCGATCGGCAAGACCTCTCGGTGTCAGTCTCTCGGAGTCCGCGAGCGCGGTTCGGC 9211
 QY 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
 DB 9212 CGCGCGCTCGGCGACCTCGGGCGCTTGAATCCGCGCGCTGGAAGCGTCCCAACGTG 9271
 QY 465 AspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSer 484
 DB 9272 GACATCTCCAAAGCACTGTGTGAACCTCATCTGCCAGCGCAACTACCGAGCCCAATGCC 9331
 QY 485 LysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLys 501
 DB 9332 AAGACATCCAGACGAGGATCGCGTGACCCAGACCATCATCACTCGCG 9382

RESULT 19
 BB043739/c 34817 bp DNA linear BCT 24-JAN-1996
 LOCUS Borrelia burgdorferi fcsmid clone 31, complete sequence.
 DEFINITION
 ACCESSION U43739
 VERSION U43739.1 GI:1165250
 KEYWORDS
 SOURCE
 ORGANISM
 Borrelia burgdorferi (Lyme disease spirochete)
 Borrelia burgdorferi
 Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
 Borrelia burgdorferi group.
 1 (bases 1 to 34817)
 Dunn, J.J., Butler-Loffredo, L., Kieleczawa, J., Medalle, J. and
 Luft, B.J.
 Direct Submission
 Submitted (21-DEC-1995) John J. Dunn, Biology Department,
 Brookhaven National Laboratory, PO Box 5000, Upton, NY 11973-5000,
 USA
 FEATURES
 Location/Qualifiers
 1..34817
 /organism="Borrelia burgdorferi"
 /mol_type="genomic DNA"
 /strain="B31"
 /db_xref="taxon:139"
 /map="approx 290-325 kbp"
 /clone="31"
 complement(<1..268)
 /note="orf39; N-terminal portion; Method: conceptual
 translation supplied by author."
 /codon_start=1
 /transl_table=11

CDS

```
/protein_id="AAA85590.1"
/db_xref="GI:1165251"
/translation="MDTFEYFVMSVGSFKARQPLSEKHKELKINQARPFELRDC
HTDSEGVNVCVNVKGVNYSLLSPFKVKNKEKVMLSLYPDKEF"
complement(295..2199)
/notes="orf37; Method: conceptual translation supplied by
author."
/codon_start=1
/transl_table=11
/protein_id="AAA85591.1"
/db_xref="GI:1165252"
/translation="MDDRAINFELCDKMSYLERSHVDLMEIADTLEALNDASL
ELSPYKELDYELLKNGSLGKGGKVIAYRNSKSPKVSLSKNDSDGEKVIS
SDGFFTRKSGKSVLFKVIKPGQGIPIKQDVIKLELHNTENFKDVIKGIENA
SGYKIDPESDESNTVMVQISNSMSVTIEFTAPINGAEIQAQDIFNIRKYG
VIDAILKVKVFDVPYVGEPELARGLEPVKGDSDYDFIVDENNRGFEYVSGI
GFRVIRGEBELARIIPFGKVDGVTVFGKVKESGODIDFILGENTFDGKYIRAKC
NGYMSVDGVLVHNIYLVKGDVGPATGDIWNGMVLVQGVLDGNIMAKSGIEVKG
LVGRCNKITGSIIFHSGVNGKGDSPKIYARGNVRKPLENTLNCGEVEVLKGVNS
SVSKKRVLCIGKSKIVGNCAKEVRAYSIGDRSCETCIDVDYDEIKDLTRF
TSLKEIKELVLTQDIVALKXNIVLITDKAEKSLKIDSPNEFVNESKILMEIKMI
EAKINLQNELENTKIEGKVSVENIACSGVKLYIKNAFYELSKDYNNVTFIEDNYIK
VMTYIPFKSR"
complement(2202..2681)
/notes="orf36; Method: conceptual translation supplied by
author."
/codon_start=1
/transl_table=11
/protein_id="AAA85592.1"
/db_xref="GI:1165253"
/translation="MFTSRLKYLLTYSALPISAVLGLFGLSFTTILVRLSLQFV
FFVIGLIIYIKVYLSNLFSTELSGNMENKQDDKSKDQFKNLDQONKSLYEN
SQNLSGDFIEBVKYKFEEDVSRGDKNKKWSFIEDNPKVADAKILMSKKE"
complement(2688..3533)
/genes="ylxH"
complement(2688..3533)
/genes="ylxH"
/notes="putative ATP-binding protein"
/codon_start=1
/transl_table=11
/product="ylxH"
/protein_id="AAA85593.1"
/db_xref="GI:1165254"
/translation="WNLIFQLMRKFGNKTFRFIAVSGKGKGSNTAIGLAKYSE
LGKVLILDAGICMANVILLGIVIPKYSIYHIAQSRDIREVITTEYNDILLACASG
TWELLDSDVDNKNFKELKIKIYEYDVIDTDSAGISQVIFSLSSDDVIVITPEP
TSITDAYGIKLSYKMKENKRLIVNRVNSKGVAKKVIDISGQFLNMDIL
GHYEDQNRSSVFKQPFVLNPNFSKASYCLDSIVATLEEVSLDNKRKRGVIGFILR
FGVGE"
complement(3588..4754)
/genes="flhP"
complement(3588..4754)
/genes="flhP"
/notes="flagella associated putative GTP-binding protein"
/codon_start=1
/transl_table=11
/product="FlhP"
/protein_id="AAA85594.1"
/db_xref="GI:1165255"
/translation="WVOYETKQPTVNEVIELIKKYGKARWMTYKTVPHGGILGLF
SKDWBSGVRYDIGNQINVEDEKRLQIKSENSIEDVLEKSLKTELAAHK
KENINPTTKIDUREDNFENYKIDINEFKREFSLSDDDVREVRVEDVLYIAK
TKGSGSIDNKKRVILVPGTGKTTIAKLAAYINGESKSLNIKITIDNRY
IGAKQIOTYQIMGPVRAIESFDKLDDEITDSKDFDILVDTICKSPQDFMLAEM
KELLNACGRDAEPHLAESTTVDKVEIFHOFSPFNKYTVFTKVDETTCVGNLISL
IYEMKXSVSVTDGQIVPHNISVABELTPIRINGRISRDDAEFIKKIKSKSY"
complement(4795..6816)
/genes="flhA"
complement(4795..6816)
/genes="flhA"
/notes="flagellar protein required for flagellar formation"
/codon_start=1
/transl_table=11
```

```
/product="FlhA"
/protein_id="AAA85595.1"
/db_xref="GI:1165256"
/translation="MGLNKKSDLLISVGLIFVAVGFLPLPAVLVDLVLITVNLVISLL
LILIVLSKRSDFSVPFTLLVMTIFGLVINSTRLLITLKGFIDQGMIRITGTIV
VGSQIQLGIVIGFIIFIIIAVQFIVITKATRVAAVAFALDAPKQKMAIDSAYS
SGNITTEBARKQNDLQSENVFYGAMDGASKFSGNVKVGFLITLINLGLLVLGITL
QGLNFNEALNNVSLTVGDGLVSQPLSLITSTGLIVTRISKNSFGGIFPDQFTNH
LGIYIVSGFLFLAFUGPFTLILMFLSLSIAFLAYSLSGIROKREIDKMKLEEQ
ASIVSKDQSPVPDLPLALGYNLAVPVDVDTKTSLLDRIVKIREIAFEGIVVP
KIRVDNMRLEPNAYSFKRGVEVGRGEIKLGFVLNVGDSIGDGLVDPGFGP
SLWVDDGREFAEKLGIVVDPDSIIATHMTelikRHSYELTQDVQNTLDFPKDY
GAIVEELKYNFSGVEIORVLQGLKEQVSRNLVITFETADFTSITKIDFIFIEKCR
QSVQRQITSGYLDLNSLNTLNPSPQMIIDSRVESNHDLISSIDPNIKTKPIYEL
FKIVNEAQEGFYPVLSSSSRPIIKVITSREIPDLVWMSVLEVPKY"
complement(6861..7979)
/genes="flhA"
complement(6861..7979)
/genes="flhA"
/notes="flagellar protein required for flagellar formation"
/codon_start=1
/transl_table=11
/product="flhA"
/protein_id="AAA85596.1"
/db_xref="GI:1165257"
/translation="MIRDEFLIKSWYIPLDFFSADDEGRTELPTDQKKQKAREGRVL
KSTEINTAVSLLLFALFFFLSYFALDLIAVFEQAIKLPEVNRMSVYTMGPAYIRS
IMGYVLFFFASLAVNFVNIIQVGFITFKSLEPRWDKISFNFSWAKNFFSAGAF
FNLFSLLKVVIIICLIYFIENNIKISLSEYTIQSISIVLVIAYKICFSSVMPFL
IQAQVDFYLFORSOYVIESLKEEVKQEKMEGDELLRSRIKERMVILSTNLRAVA
IPQADVITMPEHFAVAIKDSTMTAPKVLAKQDEIALTIKKIARENVPVLMENKL
LARALYANKVNEIIPREYWEIVSKILVRVYSITKFN"
complement(7979..8788)
/genes="flir"
complement(7979..8788)
/genes="flir"
/notes="flagellar protein required for flagellar formation"
/codon_start=1
/transl_table=11
/product="flir"
/protein_id="AAA85597.1"
/db_xref="GI:1165258"
/translation="MYWDEILNLFVLSFTILPVIRFMLKSPFSTIKICY
FNFFSILSVIVVEKIKIYPLDNMLSPALILLGEAILGLIOAFFVNIIPNVHLVG
FFNSQIQLAVNIEDVSEDSMIISQIFAYLFLLLFLSSDFLREFFVIGIHDSVLN
IRVEHVNMRNSGFVKLLMSFGFLPEKALLISFPILSLLLLFVLVLGILSKSPQIN
LLIISFTSLFGLLLIYIGFPSLAISSKRVLSLSLDSLASFLKIFSRVLK"
complement(8800..9063)
/genes="fliQ"
complement(8800..9063)
/genes="fliQ"
/notes="flagellar protein required for flagellar formation"
/codon_start=1
/transl_table=11
/product="fliQ"
/protein_id="AAA85598.1"
/db_xref="GI:1165259"
/translation="MTAGHILYLRISNIENIILSAPMLIAILVGLLISIFQAITSI
ODQTLSPKIVLIVLIVFEGPWILNKLKMQFTYMFSQLQNV"
complement(9072..9836)
/genes="fliP"
complement(9072..9836)
/genes="fliP"
/notes="flagellar protein required for flagellar formation"
/codon_start=1
/transl_table=11
/product="fliP"
/protein_id="AAA85599.1"
/db_xref="GI:1165260"
/translation="MRKCNFLFFSVTSLSPAQTSKLTQTNGLNPNFPNFSIGGSE
IAFSLQILLITITISPAFLVLSIVLDFIRRALSQSPPTQVYMGALF
LTTFTWMPFNSIYEQALPLKESKNFENFKIAPLRIFMYKMSGRHEERLFL
MNSMNDPRPKNFSEVTHVILAAFILHKLVAFLKIFLPIFLVIDIIVASVLAWMG
```

```

MIMLPVVMISLPFKLILFVMDGWTLLITSLIKSFM"
complement(9847. .10473)
/sequence="fliz"
complement(9847. .10473)
/sequence="fliz"
/notes="flagellar protein required for flagella formation"
/codon_start=1
/transl_table=11
/product="Fliz"
/protein_id="AAA85600.1"
/db_xref="GI:1165261"
/translation="NNKFLVFLVLFSENFYVANSOVSASTDLNENESLPIFER
DKANFVNDSTQPSIFINISDLKIVLFLLIAFFFLKLLIFYSKSKYEQSNLI
KAKFDYDVKNRIILNDVYIFLVSSNFTLLKREIKSEEDLKKRLSKINDS
AKKDSFQSIKKMLKKEIPSSGNDYVKLEKIEASLKQDKRLK3"
complement(10466. .10807)
/sequence="flin"
/codon_start=1
/transl_table=11
/product="Flin"
/protein_id="AAA85601.1"
/db_xref="GI:1165262"

Alignment Scores:
Pred. No.: 1.82e-22 Length: 34817
Score: 509.00 Matches: 150
Percent Similarity: 41.62% Conservative: 71
Best Local Similarity: 28.25% Mismatches: 190
Query Match: 19.68% Indels: 120
DB: 1 Gaps: 13

US-10-009-823a-1 (1-502) x BBU43739 (1-34817)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 15670 ATGATGAGCTCTTATATTCGTGTTCTGGCTTCAGAAATCATCAACAAAGATGGAT 15611
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 15610 GTTGTGGTAAACAATATCCCAATGTAATACAAATGGCTTTTAAAGAGGAGNATAAT 15551
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
Db 15550 TTTCAGATATGATATCGCAACTATTTCTGGAGCTTCTCGCCCTACTGATCTGTGT 15491
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 15490 GGGACTAATCCCAAGCAAGTTGGATTAGGCATGAATGTTGCCTCAATTGACACTATTCA 15431
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
Db 15430 ACTCAAGGAGCTTTTCAAGCACTCAAAAGCATCTGATCTTGGAGTTAGTGGCAACGGA 15371
QY 96 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Db 15370 TTTTATTATTTAAAGAGGTAAATAATTTGTTTATACAGAGCGGTGCTTTTGACGTG 15311
QY 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer 135
Db 15310 GACTCTGATGACATCTTGTAAATCTGCAAA-----GGAAATGCGAATCAA 15263
QY 136 Asn-----AsnProAsnIleLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 15262 GGTGGTGGCAGAGATTAGAGGTGAAAGGTTATAAATACAGCTTCTGATATTGAG 15203
QY 154 Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
Db 15202 GATCTGATTATTCGATTGAGATAAAGAGGAGCAAGTCTACGAAAAAATGTTACTTTT 15143
QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProThrPheAla 190
Db 15143
```

RESULT 20
AB001137/c

LOCUS AB001137 44380 bp DNA linear BCT 17-MAR-2003
 DEFINITION Borrelia burgdorferi B31 (section 23 of 70) of the complete genome.
 ACCESSION AB001137 AB0000783
 VERSION AB001137.1 GI:2688160
 KEYWORDS
 SOURCE
 ORGANISM Borrelia burgdorferi B31
 Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
 Borrelia burgdorferi group.
 REFERENCE 1 (bases 1 to 44380)
 AUTHORS Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickley, E.K., Winn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Karavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Weidman, J., Uterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.
 TITLE Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi
 JOURNAL Nature 390 (6660), 580-586 (1997)
 MEDLINE 98065943
 PUBMED 9403695
 REFERENCE 2 (bases 1 to 44380)
 AUTHORS Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickley, E.K., Winn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Karavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Uterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 Source
 1. 44380
 /organism="Borrelia burgdorferi B31"
 /mol_type="genomic DNA"
 /db_xref="taxon:224326"
 134. 1516
 /gene="BB0261"
 134. 1516
 /gene="BB0261"
 /note="hypothetical protein; identified by Glimmer; putative"
 /codon_start=1
 /transl_table=11
 /product="B. burgdorferi predicted coding region BB0261"
 /protein_id="AAC66691.1"
 /db_xref="GI:2688209"
 /translations="MLSLFTIVLSLVFVLVFLFPKIALKLGIKNGIKKDDERTR KLIERAISLKTNPGEALGILNYYNKKDHENGKIVAKLQCLIEDNPINQSEINT FRAFLSYGNLXRNFRNEALEYLRKAFIKTDDANDYLVGFALKNEMYKEALYKL KVVHFNKNKKVLLKHLGITLFLNLSYQRAVIIFNINKNIQDIEALIAVAKSLRM NOHLEALEANKIQKQDGMVYEAALLTSIHSKNKLEKLEONKEIKIVKPDLPKPT FLFLVNLGLQISLVNOKYQAEATFKVEIDPNYNIKEKLEFSLRNLENALRIYL RSKNEPKETANLIIKLINLKQIRDSKINVTQFDINFLANNQWENLIVRFV RTQDTFGELFLDFISKIKENKIGLCIAPSFKSLAKQMIEGRLLIDVEGKLAQI LKKINISKYT"
 complement (1523. .2776)
 /gene="BB0262"
 complement (1523. .2776)
 /gene="BB0262"
 /note="similar to PID:1652679 percent identity: 36.49; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC66685.1"
 /db_xref="GI:2688203"
 /translations="MIIPKKQVRRKKKFLFNSKSVNPELKDFANISNIGRRKK VFIKNFFPKKINFFKKVLSLPFYKLQINHYEYKYYSLSLRDKDFDIFSVPDYKL

VFKNALIFILFTFYINIFSYGVFLNRLSLPKDYFIDTFLYTSDDQIAQISSYL
 PESNVASGFGKKNFVLKVDHKIKFGETLSHVAARYQITSETLSFNEIKOVNRNK
 PNVIKVPNMKGIVYIVKNDSSISSASAVNPKVDILDSNLDNEVLFLGQKLFPG
 GRPKDLEKVELGETFYVQGVITSGYVERPDPFTGVSIFHNGIDIANLANTPKAS
 REGVVVTAGNAGGYKGIYIVISHSNGFQTLIAHLNSFAVKGVKVSAGVIGMGSTG
 YSTGNHLHFTIFKNGKTENPMKYLR"
 2838. 3344
 /gene="BB0263"
 2838. 3344
 /gene="BB0263"
 /note="similar to SP:P33697 GB:X52093 GB:X54933 PID:47757
 percent identity: 29.61; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="signal peptidase I (lepB-3)"
 /protein_id="AAC66684.1"
 /db_xref="GI:2688202"
 /translations="MRIFKAHYELASILAACLFLTLIKLFLSYIVKGSMTPAIF
 EKWIVNHPFAYGLRIKKHQYLLWKTPQNMVLIKDPITNKIAIKKIIPAIPGKRF
 KQIKNTICIDHLNFKIDENILKNTKKIPEDHYLVIGENKQISLSDRYGFIKIDNI
 LKIIYVF"
 complement (3370. .4845)
 /gene="BB0264"
 complement (3370. .4845)
 /gene="BB0264"
 /note="similar to GB:M96847 SP:P28608 GB:M97912 GB:S42385
 GB:X67646 percent identity: 40.88; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="heat shock protein 70 (dnaK-1)"
 /protein_id="AAC66683.1"
 /db_xref="GI:2688201"
 /translations="MFMKWIGIDLGTNTVASFYFDVSSKIIINERGITPSIVSFS
 DKDLVGSAAKQNLVNPQTFYFNKTNIGNNFYKGVGEFYKAEYISAHLASVKNR
 AEKFLDEIENAVITVPAYFSEIQRVCVEAANFAGLKAILNEPTAAIAVAFERQ
 IDGFLIYDLGGGTFDVTLMKQGDYTVLSVKQSLGCGDNFKLIEKVLDSFKNK
 YPFNLEIDFLQDRRIEIEGKNLSVMEVDITLLFLDKHLNVLKRDENSMIS
 EYDKITQLMECIAQSGVDISSVKIILSGSGSTRIPLIEKVLKESFPFSAIIDLALNQ
 DEVAIGAGIHAFSLSDNSVIRKFDVTPYSLGLEIKDNEFFTLIKENTALPIRSKL
 FTTNDYQDEIEIHLIKGEYKASLANSIGRFSFGNIQKALGKIPKIEVFLTLNESGI
 LSVNAKLETNNSNFTIEIKITSSSDNVAKESLSNTFTSIID"
 complement (4817. .5338)
 /gene="BB0265"
 complement (4817. .5338)
 /gene="BB0265"
 /note="hypothetical protein; identified by Glimmer; putative"
 /codon_start=1
 /transl_table=11
 /product="B. burgdorferi predicted coding region BB0265"
 /protein_id="AAC66690.1"
 /db_xref="GI:2688208"
 /translations="MFIFYLIIILIFIFYVHYVINLKSNILKRFKSEVDKTIIE
 INQATDNINIIIEIKESLNRIKEVDREIIEQLDRLGNSSIFGNSSPSSGSKSD
 GIKYNSLDYSGMPTIEKNIIEKESVNRNRIILHEQGMFPAIAKFKLDLGEVELIIS
 IHRGCVHEKMDRY"
 complement (5358. .5666)
 /gene="BB0266"
 complement (5358. .5666)
 /gene="BB0266"
 /note="similar to PID:1165251 GB:AE000783 percent
 identity: 100.00; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC66682.1"
 /db_xref="GI:2688200"
 /translations="MYMDTEVFMKSVYGVSKARQFLSEKHKRLKINONARFRFLR
 DCHTDSVGVNVCVNVKNVGDNYLLSFPVKKNKEKVMLSLYPDKFDPKIGKNVNIKR
 "

gene	complement (5687. .7591)	41	PheGlnAspLeuPheSerGlnAspLeuAlaileGlySer	-----ThrGly 55
CDS	/gene="BB0267"	20340	TTTCAAGATATGATATCGCAATCTATTTCTGGAGCTTCTCGCCCTACTGATGCTCGTGGT	20881
	/note="similar to PID:1165252 GB:AE000783 percent identity: 100.00; identified by sequence similarity; putative"	56	SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe	75
	/codon_start=1	20880	GGGACTAATCCCAAGCAAGTTGGATTAGGCATGAATGTTCCCTCAATTCACATATTCAC	20821
	/transl_table=11	76	ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaileGlyGlySer	95
	/product="conserved hypothetical protein"	20820	ACTCAAGGAGCTTTTCAAGACACTCAAAAAGCATCTGATCTTGGAGTTAGTGGCAACGGA	20761
	/db_xref="GI:2688199"	96	PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe	115
	/translation="MDDRINFSELCDKMSRYLERSHVDLMEIADTLERALNDASL	20760	TTTTTTTATTTAAAGAAAGGTAAATAATTTCTTTTATACAGAGCCGGTCTTTGACGTG	20701
	ELSVPEKLDYDILLKNGILGYGKKWKIVAYRNSKPKVSLDSDSGSEKIVIS	116	ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer	135
	SGEPFIRKSSGVFLKVIKSPKGQIPIKYQDVIKELHNSNIENFNKDIKVGIVENA	20700	GACTCTGATCGACATCTTGTAAATCTCGCAAT	20653
	SCYKIKIADFSDDPSNTVMQIISNSMSVTIIFTAPGINGAEIFAODIENILKRYG	136	Asn-----AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAsnPheAsn	153
	VDIALKDKVAKFYDVPYGVPELARGLEPKGDSYIDFIVDENNRFCGEYVSSI	20652	GTTTGGATGCAAGAGATTTAGAAGGTGAAAAGGTTATAATACAGCTTCTGATATTGAG	20593
	GRNVIEGELARIIPFGGVGDYTVFGKLVKSGQDIDFILCENTFRDGYKIRAKC	154	Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla	170
	LYRNCNLKTDGSIIFHSYGVGKDSKIYARGNVRKAFLENTLNGCGEVEVLRTGVS	20592	GATCTGATATTCCGATTGGAGATTAAGAGGGAGCAAGTCTACGAAAATGTACTTTT	20533
	SVSSKRVLCIKGKSKIYNSVCAKEVRAYSIGDRSCETCIDVGYDPEIKDLITRF	171	ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla	190
	TSHEKIKRLVLTKDIIKVALKNIVLITDKAEKSLKIDSNEFYNESKILRMEIKMI	20532	GCTTGTAAATCTTGATAAGAGATTCCTTAATCAAGAGGTGCAATCTCTGAGATATT	20473
	EAKRLNLOELENTEKIEGVSVENIACSGVKLIYKNAFYELSKDYNVNTFFIEDDNYIK	191	LeuLeuGluSerThrLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyr	210
	VMTYIPFKSR"	20472	GCACCGCGAACTTGGGTTGTCAATAATCA	20443
	complement (7594. .8073)	211	AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPhe	230
	/gene="BB0268"	20442	-----TTGTATGACAGTTTGGAAAT	20410
	/note="similar to PID:1165253 GB:AE000783 percent identity: 100.00; identified by sequence similarity; putative"	231	AspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnPro	250
	/codon_start=1	20410	-----	20410
	/transl_table=11	251	SerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThr	270
	/product="conserved hypothetical protein"	20410	-----	20410
	/protein_id="AAC66680.1"	271	MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer	290
	/db_xref="GI:2688198"	20409	-----GAGCTTAGA	20401
	/translation="WFISRELKVIILTSVSALFISAVLGLFGLSFFTLVRSLLQFV	291	AlaThrLysAspLeuAsnAla-----TrpGlnProAlaProLeuValAsnGlyLeu	307
	FFVGLLEIYFKYLSNLFSTELSGWENKPDQDKSDDFKFNLFQNKSLYEN	20400	GTTGTAAAAGATTTAAATACGCTTAATTTATGGAATGCAACAGATTATAAATGGT---	20344
	SONISSGGDFIEVKYIKFETEDVRGSDKXKMSFIENDPKRVADAIKILMSKE	308	ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe---	326
	complement (8080. .8967)	20343	---GAGCAAAATTCAAATTTT---	20311
	/gene="BB0269"	327	-----GlyIleLysSerGlnGlnAsnMetTrpAlaGly 337	
	/note="similar to PID:1165254 percent identity: 99.64; identified by sequence similarity; putative"	20310	AATGAAGGAGCATTTGGCGTCTTTAAATGGTCAACAGCGGCAAAAGAGAGATATTCCTCAA	20251
	/codon_start=1	338	AlaProAlaSerAlaAlaAlaileGlyThrAspIleGlyLysLeuProSerMetMetPro	357
	/transl_table=11	20250	ATTCCTATAACATTTAATGTTTGGTGCAATAGTAGTGAAGTTGGTGGACAGCAAACT	20191
	/product="ATP-binding protein (ylxH-1)"	358	IleGlnThrSerSerGly-----AsnSerThrAlaArgAsnGly 370	
	/protein_id="AAC66679.1"	20190	GTAATTTGAAATTCGGAACACAGTTGGAAGTTTACATGATTCATTAATCTACTGCTGAT	20131
	/db_xref="GI:2688197"	371	SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal	390
	/translation="VEDQAQSLRDMMLNGKFNFSVDEKVNQSKTRFIATVSSGKGVG			
	KSNIAIGLAKYSELGKVLIIADIGMANVNIILGVIPKYSIYHMIQSRDIRVIT			
	KTEYNIDLAGAGTMELLDLDSDVNRKFIKELKIYEDIVVIDTISAGISKQVLSFL			
	LSDDVIVITPTPTSITDAYGIILKIVLSYKMYENLKNLRILIVNRVANVSEAGVAKVI			

Alignment Scores:

Pred. No.:	2, 43e-22	Length:	44380
Score:	509.00	Matches:	150
Percent Similarity:	41.62%	Conservative:	71
Best Local Similarity:	28.25%	Mismatches:	190
Query Match:	19.68%	Indels:	120
DB:	1	Gaps:	13

US-10-009-823A-1 (1-502) x AE001137 (1-44380)

Qy	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db	21060	ATGATGAGGCTTTATATTCTGTTCTTGGCTTCAGAAATCATCAAAAGAAATGGAT 21001
Qy	21	ThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db	21000	GTTGTTGGTAACATATCGCCATGTAATACATTAATGAGTTTAAAGGAGAGATTAAT 20941

```

Db 20130 TCTAGTACCAAGGCTATTATTCAAGATGGATATGCGATGGATATATGGAATAATAT 20071
Qy 391 ThrileThrSerGluGlyLysLeuGlnGlyLysSerAsnSerGlnValValAspPhe 410
Db 20070 GAAATTGATCAAAATGGTGTATAGTTGGCATTTATTCAATGGCATAAGACAGATCTT 20011
Qy 411 TyrAsnIleProLeuAlaPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsn 430
Db 20010 GCGAAGATGCTCTGCTCTTTTATGAATCCCGAGGACTTGCATAATCAGCGCATACT 19951
Qy 431 HisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
Db 19950 AATTTGTAGAACACCAATTCAGTCAAGTTAGATAGCGAACTGACCTGCTGGA 19891
Qy 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
Db 19890 CTTGGTGATATTAGATCTGGTGTATTTAGAAATGCCCAATGTTGATCTTGCAGAGCAATTT 19831
Qy 471 ValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
Db 19830 ACAGATATGATAGTACCCAAAGAGGATTTACGCAATGCCAARACCATTAACCATCTCT 19771
Qy 491 AspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 19770 GATCAATTATTACAAGAACTGTGAAGATGAAA 19738

RESULT 21
LOCUS BORFLGE 1499 bp DNA linear BCT 06-FEB-1996
DEFINITION Borrelia burgdorferi flagellar hook protein (flgE) gene, complete
cds.
ACCESSION L43849
VERSION 1 GI:1184295
KEYWORDS flagellar hook protein; flgE gene.
SOURCE Borrelia burgdorferi (lyme disease spirochete)
ORGANISM Borrelia burgdorferi
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
Borrelia burgdorferi group.
REFERENCE 1 (bases 1 to 1499)
Ojalini, C., Davidson, B. E., Saint Girons, I. and Old, I. G.
Conservation of gene arrangement and an unusual organization of
rRNA genes in the linear chromosomes of the Lyme disease
spirochaetes Borrelia burgdorferi, B. garinii and B. afzelii
Microbiology 140 (Pt 11), 2931-2940 (1994)
MEDLINE 95111614
PUBMED 7812434
REFERENCE 2 (bases 1 to 1499)
Old, I. G.
Unpublished (1996)
On Feb 11, 1996 this sequence version replaced gi:903702.
Original source text: Borrelia burgdorferi (individual isolate
HB19) (clone: PAD12B) (clone library: HB19 Cosmid) DNA.
FEATURES
source
1..1499
/organism="Borrelia burgdorferi"
/mol_type="genomic DNA"
/isolate="HB19"
/db_xref="taxon:139"
/clone="PAD12B"
/clone_lib="HB19 Cosmid"
115..1443
/gene="flgE"
115..1443
/gene="flgE"
/codon_start=1
/transl_table=11
/product="flagellar hook protein"
/protein_id="AAA87351.1"
/db_xref="GI:1184296"
/translation="MMRSLYSGVSGQLNQHQTIMDVVGNNIANNVTIGPKGVNFQDM
ISQISGARPTDARGTNPQVGLGNVASIDTHTTQAFQSTQASDLGSGNGFF
ILKEGKNLYFRAGAFYVDSRHLVNPANGMRIQGMWARDLEGEKVINTASDIEDLI

```

```

PIGKEGAKSTNTTFACNLDKRLPLIQEGASPADIAKRTWVNVKSLYDSFGNVSVLE
LRVKDINTPLNWTATVLINGEONSFTLGFNEGALASLNGQGGKDLIQPITFN
VLGVGEVGEQOVTNKLGTGVSYSYDSITQFADSSSTKAIQDQVGMGMENYEDQ
NGVTGVYSNGIRDLGKIALASPMFPGGLAKSGDINFTVSTNSGVQRIGETGLAGLG
DIRSGVLEMANVDLAEOFTDMIVTQRFQANAKTITTSDDLQELVRLKN"

ORIGIN

Alignment Scores:
Pred. No.: 1,08e-23 Length: 1499
Score: 503.00 Matches: 149
Percent Similarity: 41.62% Conservative: 72
Best Local Similarity: 28.06% Mismatches: 190
Query Match: 19.45% Indels: 120
DB: 1 Gaps: 13

US-10-009-823A-1 (1-502) x BORFLGE (1-1499)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 115 ATGATGAGGTCTTTATATCTCTGGTGTTCCTGGCTTCAGAAATCATCAACAATAATGGAT 174
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 175 GTTGTGTTGTAACAATATCGCCAATGTAAATACAAATTCGCTTTAAAAAGGAAGAGTAAT 234
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
Db 235 TTTCAAGATATGATATCGCAATCTATTTCTGGAGCTTCCTCGCCCTACTGATGTCGTGGT 294
Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 295 GGGACTAATCCCAAGCAAGTTGGATTAGGCAATGTTGCCTCAATTCACATTTTAC 354
Qy 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGly 95
Db 355 ACTCAAGAGCTTTTCAAGCACTCAAAAGCATCTGATCTTGGGGTAGTGGCAACGGA 414
Qy 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
Db 415 TTTTATTATTTAAAGAGGTTAAATTTGTTTATACAGAGCGGTGCTTTTACGTG 474
Qy 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer 135
Db 475 GACTCTGATGACATCTTGTAAATCCTGCAAT-----GGAATCGCAATTCAA 522
Qy 136 Asn-----AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 523 GGTGGATGTCAGAGAGATTTAGAGAGTGAAAGGTTATAAATACAGCTTCTGATATTGAG 582
Qy 154 Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
Db 593 GATCTGATATTATCCGATTGGAGATAAAGAGGAGCAAAAGTCTACTAAAAATATACCTTT 642
Qy 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla 190
Db 643 GCTTGTAACTCTTGATAAGAGATTGCCCTTAATTCAGAGAGTGGAGTCTCGCAGATATT 702
Qy 191 LeuLeuGluSerThrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyr 210
Db 703 GCACGCGGAACTTTGGGTGTCAATAAATCA----- 732
Qy 211 AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPhe 230
Db 733 -----TTGTATGACAGTTTGGAAAT-----GTTAGTGTCTT--- 765
Qy 231 AspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnPro 250
Db 765 ----- 765
Qy 251 SerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThr 270
Db 765 ----- 765

```

271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
766 -----GAGCTTAG-----
291 AlaThrLysAspLeuAsnAla-----TrpGlnProAlaProLeuValAsnGlyLeu 307
775 GTTGTAAAGATTAAATACGCTAATTATGCAATGCAACAGTATTATAAATGCT--- 831
308 ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe--- 326
832 ---GACCAAAATCAATTTT-----ACATTTGGGTTTGC 864
327 -----GlyIleLysSerGlnGlnAsnMetTrpAlaGly 337
865 AATGAAGAGCAATGCGCTTTTAAATGCTACACAGGCAAAAGAGATTTCTTCAA 924
338 AlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetPro 357
925 AITCTTATACATTAAATGTTTGGTGCAATGTAAGTAGGTGAAGTTGCTGAGCAAACT 984
358 IleGlnThrSerSerGly-----AsnSerThrAlaArgAsnGly 370
985 GTAAATTTGAAATTTGGAACAGTTGGAAGTTACACTGATTCATTAATCACTGCTGAT 1044
371 SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal 390
1045 TCTAGTAGCACAAGAGCTATTATCAAGATGATGATGCGATGCGATATGGAATATAT 1104
391 ThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPhe 410
1105 GAAATTTGATCAAAATGCTGTATAGTTGCAATTTATCAATGGCATAAGACGATCTT 1164
411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsn 430
1165 GCGAAGATTGCTCTCTCTCTTTATGATCCGAGAGACTTGCATAATCAGCGCATACT 1224
431 HisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
1225 AATTTGTAGAACACAGCAATTCGGTCAAGTAGATAGGCAACTGCTGCTGCA 1284
451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
1285 CTTGGTGATATTAGATCTGCTGTTTATGAAATGCGCAATGTTGATCTTCAGACAAATT 1344
471 ValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
1345 ACAGATATGATAGTACCCCAAGAGAGATTTCAGGCAATGCCAAATCCCAATACCACTTCT 1404
491 AspThrMetLeuGlnLysAlaLeuGluLeuLys 501
1405 GATCAATTATTACAGAACTTGTAGATTGAAA 1437

RESULT 22
AL646078 203050 bp DNA linear BCT 02-SEP-2002
LOCUS Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
DEFINITION segment 3/11.
ACCESSION AL646078 AL646053
VERSION AL646078.1 GI:17430778
KEYWORDS
SOURCE Ralstonia solanacearum
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE 1
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Ariat,M., Billault,A., Brottier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choize,N., Claudel-Renard,C., Cunac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
Siguer,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Nature 415 (6871), 497-502 (2002)

MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

21681879
11823852
2 (bases 1 to 203050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMG CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1..203050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
complement(147..1499)
/gene="RSP0310"
/note="synonym: RS05460"
complement(147..1499)
/gene="RSP0310"
/functions="cell processes; transport of small molecules"
/note="product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PROBABLE MACROLIDE-EFFLUX TRANSMEMBRANE PROTEIN"
/protein_id="CAD17461.1"
/db_xref="GI:17430779"
/db_xref="GOA:Q8XT07"
/db_xref="TrEMBL:Q8XT07"
/translation="MNAPRKNTDRAEAALLRNASTTHEPCWIPDAAAAGLHPGR
WRAPFAIFGQALSLIGSALTOTVLLWLTDTGVSALATAGLALLQAVLSPLG
GTFADRSRLMIVADGIALCMLLLALFLTGRIALWYVYVMAIRSAQAQFAPA
AASVTMLVPGGLVRAAGLNLQSLITLVAAPLGALAIGVMPDGLWALGIDVATALL
GIAPILCCRIPQACVSSGQKTLWPAFEGVDLVWKTPTGLRWLVGLVLTIMPTF
TLVPLVTHFGGAAGQVAMLEGLSGIGMAGVLLVAMAPRRQVAVILGCFAMSCVA
LALTAPAGSLFGVAVAWVIGITTFVGNAPLTALQITVPNHLQGRVLSLNTTNG
LAAPVGLALTPLGEVIGVRAIPALGALGAAASLAGFLSPSLMELDRPFRSEARTD
"

complement(1531..2961)
/gene="oprX"
/note="synonyms: RS05459, RSP0311"
complement(1531..2961)
/gene="oprW"
/functions="cell processes; transport of small molecules"
/note="product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="PROBABLE OUTER MEMBRANE CHANNEL LIPOPROTEIN"
/protein_id="CAD17462.1"
/db_xref="GI:17430780"
/db_xref="GOA:Q8XT06"
/db_xref="TrEMBL:Q8XT06"
/translation="WELRIILPLVIALSACSLIPALIKPALVPATFPABSPETQAH
ADLGRWTFGDRRLQRLIGLITLANNRDLRLALNVEVYVYQYSIQRAARLPGLDQGS

ASRQRTAANDGAPPAALANSEQYGVNVLNAFEIDLFGVRSLSDAAFYARLASQDG
RRAQISLVGADAVFAERLAQEQOLAEHTLADWRQSLDLARRLKTAQNSDLDDVA
QAEQVANOADLEARTLALAQAGNSLORLVGAGLPADLPAPLSLETQPVVTLQAGL
PSLEIPRDIROAEHNLAAANDIGAAARAFPRLSLTASLGYSAPAGSLFDGGRH
VWFSPOITOPILQGGRLAELRLAELRKSSAAEYERTIQTAFREVAUGLAGRATYG
RQIEAQIRAVASRENLNRYRAGVBEGRLELDSQRLYASRQALLDLRQGAFSN
AVALYKALGGGLAETDVATDRVARK
/gene="mexD"
/note="synonyms: RS05457, RSp0312"
/complement(2966..6115)
/gene="mexD"
/note="synonyms: RS05457, RSp0312"
/complement(2966..6115)
/gene="mexD"
/functions="cell processes; transport of small molecules"
/note="Product confidence: probable
Gene name confidence: putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM TRANSMEMBRANE
PROTEIN"
/protein_id="CAD17463.1"
/db_xref="GI:17430781"
/db_xref="GOA:Q8XT05"
/db_xref="TrEMBL:Q8XT05"
/translation="MPQFFIRRVFAWVIALFILGAIAPKLPRIAPVSPVPTVN
LSATPGATQTNWDGVLISIRELSVNKLIFESSSDTSASIVATFKECTDPAM
AQVDQNKLAIEPRLPQVVRQGLTVSAGSSGLMIIGLQSDGRFDELTALDYMAR
NVVEELRIEGRVQLVGAEOAMRIWDPAKLIAVGLSNDLSAAISOONAQIAPGS
VGSAPALPAQRTVPLTVQGLTTPQFAIIVLRANAGSKVVMGVARVELGAQSYG
FINREKQAATAAVALQAPANAVTAEGVQARMELSRIMPAGMSISIPNTAFVK
ISIEKVQTLIEAMVFLVFLYFLQNVRYTLIPALVAPILGTLFTIMLAGSFINV
LTFMGVLAIGTIVDAIVVVENRIMAEQSLPRDITKMKETIGAVIGITVLIT
AVTIPMALAGSVGVIVQOFTSMNVISVFSAPLALTLPALCATLTKPDPFAHEKR
GVFGVNRDRITQGVAAVGRGLVTRGRVLMVFGAIAAVLALAFQPLPSFLPEED
QGVFMTTLPADATAEITLDVYKAFERHVASASIQSNMSIVGFGSGSGSNAMAF
TMKMDPARATQAEMLAQQAQAGATEGVNLSLIPALDELGTSSGFTLRQDRA
NQDAAKAAETKLELAAQSRISGVIPDLPAGTSIRLIDRQAEALGVAFTSIS
ETLSATGSLVYNDFFNAGMOQVITQADAPARMQIDDLKLYVRNAGGVPLSEW
RPVMSPTQLQVRFQGPAPARIIGNAPASGAAMERLAAQLPPGFAIAWTGQS
LQERQSAQAPMLMSLVVLVLAALYESWALPLSVMLVPLGLGVALVLRGM
PNDVFFKVGMITVIGLSAKNAIILVFEAKQLREQGGLVESAVTAARLRILPILMTSL
AFGLGVVPLMIANGASNETQHAITGVFGVGSWATVLAIFFPVFPVFWMLQERPEA
WRASRQAPPPPPHQFQEG"
/complement(6136..7293)
/gene="mexC"
/note="synonyms: RS05456, RSp0313"
/complement(6136..7293)
/gene="mexC"
/function="cell processes; protection responses;
drug/analog sensitivity"
/note="Product confidence: probable
Gene name confidence: putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM PROTEIN"
/protein_id="CAD17464.1"
/db_xref="GI:17430782"
/db_xref="GOA:Q8XT04"
/db_xref="TrEMBL:Q8XT04"
/translation="MKIERLRHGGVHVLLAAMLAGGSEBQSPAPPAEVAFTTVAP
AHLEETDLGRVAAVRVAREIQSVGIQVRRLFEGQTEVRAGQELFQINPAFPADA
DTAAALRLRAALAEARVQTTQLPLVEADAVSRQYDDAVSQDDAAQAEVQERAT
LAARQLKPAATEALPIPRIDQALVTEGALVSSGSDSNPMARIQQIDVYVVRFPAS
RQARQLTLATQKTGADGLPFAVLRNGEPYKAGHILFSGINTVDAGTGDDVLLRVLV
NNPQROLPLGMFVRVAPHARYTDAITVPOQAVRVVGQPOVMTLIDANSHARKLPVEL
GELADRRYRIAGLSAQKIVSGMERLSLDSGAHVAHDKWSPEPVPAGPAR"

7392..8162
/gene="RSp0314"
/note="synonym: RS05454"
7392..8162
/gene="RSp0314"
/function="miscellaneous; not classified regulator"
/note="Product confidence: probable
Gene name confidence: hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE TWO COMPONENT RESPONSE REGULATOR
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17465.1"
/db_xref="GI:17430783"
/db_xref="GOA:Q8XT03"
/db_xref="TrEMBL:Q8XT03"
/translation="MEINORGAAYVPMPPPSGAQSAQALLIIAEDFEIADILAA
YLAESGURSVAADGREALHLSKLPDLVLLDVQMPHVGWKSLSIRHRGDTPIVM
LTALDQDIDKLMGLRIQADYVVKPFPNPAVWARTQAVLRSMGSTRGRDVRVAP
FQIDLNHEATVQVQKORHALALTLEFKLLAQLARAPKRVFSRAELLAAACLPEDGTL
ERTVDSHISIKLRKLEDLGVNMPVGVGVYRLWSLE"
8159..9241
/gene="RSp0315"
/note="synonym: RS05453"
8159..9241
/gene="RSp0315"
/EC_number="2.7.-."
/function="miscellaneous; not classified regulator"
/note="Product confidence: probable
Gene name confidence: hypothetical
predicted by Codon usage

Alignment Scores:

Pred. No.:	4,2e-21	Length:	203050
Score:	502.00	Matches:	137
Percent Similarity:	44.40%	Conservative:	85
Best Local Similarity:	27.40%	Mismatches:	168
Query Match:	19.41%	Indels:	110
DB:	1	Gaps:	15

US-10-009-823A-1 (1-502) x AL646078 (1-203050)

QY	8	GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla	27
Db	44995	GGCTGAGCGGTCTCGACCGCGCTCGAAGAACCTCGAGTCGCGCAACGTCGCC	45054
QY	28	AsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGln	47
Db	45055	AATGCCACACGGTCGGCTACAGAGAGTCGACGCCCGAGTTCGGCAGCGTATGCGCGT	45114
QY	48	AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln	67
Db	45115	TCGCTG---GTCGCGCGCTCC-----GACAAACAGATCGCCCGAGGCGTGGAC	45159
QY	68	ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr	87
Db	45160	GTGACCAAGGTGTCGAGTCGTCCTTCACGCGAGGCAACGTCGACCGCGCAACCCGCTG	45219
QY	88	AspLeuAlaIleGlyLysGlyPhePheGlnVal-----ThrLeuGluAspLysVal	105
Db	45220	GACATCGCATCAACGCGCACCGGCTTCTACGGATGTCGACGCGCTCGAGCGCGAGTG	45279
QY	106	HisThrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSer	125
Db	45280	TCGTACACCCCAACGCGCAGTTCACACCGACAGAACCGGTACATCATCTCGCCACC	45339
QY	126	GlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeu	145
Db	45340	GGCCAGAACCTGACCGCGCTACGGCGTGGATCGACCGCGCAAGATCAACACCGCGCTG	45399

QY 146 GluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerPro---AlaLysThr 164
Db 45400 ACCAACTCTGATGATCCGGTGAACGAT-----CTGCGCGCGTGGCGACACACC 45447
QY 165 SerThrAlaLeuAsnAlaValAlaLeuGlyAspSerThrAspLysThrGlnSerGlu 184
Db 45448 AACACCGCGTTCTCC-----ATCAACCTC---GATCGCGCGGACGCGTGGCGAC--- 45495
QY 185 AlaAsnProTyPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheLysSer 204
Db 45496 -----TyrSerTyAlaGlnProMetArgValTyAspGlnGlnGly 45510
QY 205 ThrSerAsn-----TyrSerTyAlaGlnProMetArgValTyAspGlnGlnGly 221
Db 45511 GCCACCACTCGGCCACGCTCAACCACTCGGTGTCGACAGCGTGTACAGCGCGACGGC 45570
QY 222 AsnSerHisAspIleThrValTyPheAspGlyAlaProSerSerThrGlySerLysThr 241
Db 45571 ACCTGCGACATGTCACCACTATTAC-----GTGCGCGCGCGCGCGTGGGATGTG 45624
QY 242 PheGluTyLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAsp 261
Db 45625 TACTCGACGTGGACCGCGCAATCGACCGCGCGGCAACCGGTCACTGCGTGC 45675
QY 262 SerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn 281
Db 45676 -----CTGACCTTCAACGACGACGCGGTGCTG----- 45702
QY 282 MetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAla 301
Db 45703 -----ACCTGTCGCGGACGACGAGTGGCGTGGCGCTTC----- 45735
QY 302 ProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnPro 321
Db 45736 -----CGCGGCGATGAGCATC 45750
QY 322 LeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProLaser 341
Db 45751 GCCAATCATGACTTC----- 45765
QY 342 AlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSer 361
Db 45766 -----ACCGGACCC----- 45774
QY 362 SerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTySerGlnAspGly 381
Db 45775 -----ACCGAGTACGCGCGCGGCTTCAACGACACACACCGCTCTCGCAGGACGCG 45822
QY 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 45823 TACGCCACCGCGCGCTGCGCCAGCTACTCGTGGGACCGGATGGACCATCAACGCGCGCG 45882
QY 402 TyrSerAsnSerGlnValValAspPheTyAsnIleProLeuAlaArgPheThrSerGlu 421
Db 45883 TACTCCACCGCGCGCGCGCTGCGCCAGATCGCCATGACCACTTCAAGCGCGCG 45942
QY 422 AspGlyLeuArgArgGluGlyAsnAsnHisTySerAlaThrLeuAspSerGlyGlyPro 441
Db 45943 GACGCGGTGCAATATCGCGCGCAACCACTGCGGTGGAGACGCGCGAGTTCGCGCGCGCG 46002
QY 442 GluPheGlyLeuProGlyThrSerAsnTyGlyLysLeuSerValAsnGlnLeuGluThr 461
Db 46003 CAGATGGGACCGCGCGCGCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46062
QY 462 SerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGln 481
Db 46063 TCCACGCTGACCTGACGCGCGGAGCTGCTCAACATGATGCTGCGCGAGCGCTGCTGACAG 46122
QY 482 MetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 46123 GCCAAGCGGACGACCATCAAGACCGGAGACCATCTCTCAGACCTCTGTGTGAGTGGCG 46182

RESULT 23
STFLGE
LOCUS 1320 bp DNA linear BCT 08-MAR-2000
DEFINITION Salmonella typhimurium flgE gene for flagellar hook protein.
ACCESSION X51737
VERSION X51737.1 GI:47671
KEYWORDS flagellar hook protein; flagellar protein; flgD gene; flgE gene;
flgF gene;
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 1320)
Homma, M., DeRosier, D.J. and Macnab, R.M.
Flagellar hook and hook-associated proteins of Salmonella
typhimurium and their relationship to other axial components of the
flagellum
J. Mol. Biol. 213 (4), 819-832 (1990)
AUTHORS J. Mol. Biol. 213 (4), 819-832 (1990)
TITLE
MEDLINE 90294298
PUBMED 2193164
REFERENCE 2 (bases 1 to 1320)
MacNab, R.M.
Direct Submission
Submitted (16-FEB-1990) Macnab R.M., Yale University, Dept of
Molecular Biophysics and Biochemistry, Box 6666, 260 Whitney
Avenue, New Haven CT 06511, U S A
COMMENT See <X52094> for flgF gene.
FEATURES
Location/Qualifiers
1..1320
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
/map="23min"
<1..34
/codon_start=2
/transl_table=11
/product="flgD gene product (11 AA)"
/protein_id="CAA3602.1"
/db_xref="GI:809769"
/db_xref="GOA:P16321"
/db_xref="Swiss-Prot:P16321"
/translation="TTLDEVQRQII"
61..1272
/codon_start=1
/transl_table=11
/product="flgE protein (AA 1-403)"
/protein_id="CAA3602.1"
/db_xref="GI:47673"
/db_xref="GOA:P16322"
/db_xref="Swiss-Prot:P16322"
/translation="MSESOAVSGLNRAATLDVIGNNIANSATYGFKSGTASFADMEF
GSKVLGVKAGITQDFIDGITNTGKLDVAISQGFRLVDSNGSYFYSRNGQFKL
DENRLVNMQGMQITGYPATGPTIQGANPAPITIPNTLMAKSTITASMQLNLS
TDPVSKTPFSVSDADSYNKKGTVTVDYDQGNADHMNVYFKTKDNEMWVYTHDSDP
AATAPTASTTLKFNENGILSEGGTVNITGTNGATATFSLFLNSMQNTGANNI
VATNONGYKPDLYSVQINNDGTGVGNSYNEQVQLQIVLANFANNEGLASQGDNVW
AATCAAGVALLTAGSGNFGKLTNGALEASNDVLSKELVNNIVAQRYQNSAQTIKTQ
DQILTLVNLK"
1293..>1320
/codon_start=1
/transl_table=11
/product="flgF gene product (AA 1-10)"
/protein_id="CAA3603.1"
/db_xref="GI:581759"
/db_xref="GOA:P16323"
/db_xref="Swiss-Prot:P16323"
/translation="MDHALITYAM"

Alignment Scores:
Pred. No.: 1,96e-23 Length: 1320
Score: 498.00 Matches: 144
Percent Similarity: 43.91% Conservative: 76

Best Local Similarity: 28.74%		Mismatches: 167
Query Match: 19.26%		Indels: 114
DB: 1		Gaps: 14
US-10-009-823A-1 (1-502) x STFLGE (1-1320)		
Qy	10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla	29
Db	82 AGCGGCTGAACGGTGGCGCCACCAACCTTGAATTCGGTAAATACATCGCAACTCC	141
Qy	30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu	49
Db	142 GCCACCTATGGCTTAAGTCGGTACGGCATCTTCCGATATGTTGCC	192
Qy	50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly	69
Db	193 -----GGTTCC-----AAAGTGGGGTGGCGGTAAAGTGGCG	225
Qy	70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu	89
Db	226 GGGATTACCCAGGATTTTACCGAGGTACACACGACACCGGGCGGGTGGATGTC	285
Qy	90 AlaIleGlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr	108
Db	286 GCGATTAGCCAGAACGGTCTTTCGGCTGGTAGACAGCAACGGTTCGGTGTCTATAGC	345
Qy	109 ArgAlaGlyAsnAspPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr	128
Db	346 CGCAACGGCGAGTTCAAATGAGACGACCGTAACCTGGTCAATATGACAGGGAGTCAG	405
Qy	129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle	148
Db	406 TTGACCGGTCTCCGGCCACCGGTAGCCCGCCGACCATTCAGCAGCGGGCGAATCCTGCG	465
Qy	149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu	168
Db	466 CCGATCACCAATCCGACACCGCTGATGGCG-----GCCGAATCCGACCAACCCGCG	516
Qy	169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr	188
Db	517 TCAATGAGATCACCTG---AACTCAACGACCCCTGTACCTCTAAACG-----	564
Qy	189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProProIleSerThrSer	206
Db	565 -----CCCTTTAGCGGTGATGATGCG	585
Qy	207 ---AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp	225
Db	586 GATTTCGTATACAAAAGGACCGTACCGTTTATGACGACCGGGTAAATGCCCATGAC	545
Qy	226 IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu	245
Db	646 ATGAACGTCTATT---GTGAAAACCAACAGATAATGAATGGGCC	687
Qy	246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerClyThrAspSerAlaGlyLeu	265
Db	688 GTGTACACCATGACGACGATCTCTGAGCCACTGCGCCCAACACGCGG-----	738
Qy	266 LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys	280
Db	739 -----TCCACTAGCTGAATTCATGAAGAACCGGATTCGTGAGTCTGGCGTACGGTG	792
Qy	281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro	300
Db	793 AACATCACCCCGGTACGATTAATGGCGGACAGCG-----	828
Qy	301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln	320
Db	829 -----GCCACCTCTCTCCCTCAGCTTCCTTAACCTCC-----	858
Qy	321 ProLeuThrLeuAspPheGlyIleLysSerGlnAsnMetTrpAlaGlyAlaProAla	340
Db	859 -----ATGCAGCAGAAC-----	870
Qy	341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr	360
Db	871 -----ACCT-----	873
Qy	361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp	380
Db	874 GGGCTAATAACATCGTCGCCCAAT-----CAAAC	906
Qy	381 GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly	400
Db	907 GGCTATAAGCGCGGACCTGGTACGATTAACAGCAACGACGCGACCGTGGTGGC	966
Qy	401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer	420
Db	967 AACTACTCAACGACGACGAGCAGGTGCTGGGCGAGATTGTGTGGTAACTTCGCCAAC	1026
Qy	421 GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly	440
Db	1027 AACGAAGTCTGGCATCCAGGGCGATAACGTCTGGCGCGGACGACGAGGCTTCGGGGTT	1086
Qy	441 ProGluPheGlyLeuProGlyThrSerAsnTyrClyLysLeuSerValAsnGlnLeuGlu	460
Db	1087 CGCTGTCTGGGACTGCGGTAACCTGTAAGCTGACAAACGCGCGCTCGAA	1146
Qy	461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe	480
Db	1147 GCCTCTAACGTGGATTTCAGTAAGAGCTGGTGAATATGATCGTCGCGCAGCGTAAC	1206
Qy	481 GlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu	500
Db	1207 CAGTCGAATGGCAGACCATCAAAACCCAGACGATCTCAATACGCTGGTAACTCG	1266
Qy	501 Lys	501
Db	1267 CGC	1269
RESULT 24		
LOCUS		AE008751
DEFINITION		Salmonella typhimurium LT2, section 55 of 220 of the complete genome
ACCESSION		AE008751 AE006468
VERSION		AE008751.1 Gi:16419687
KEYWORDS		Salmonella typhimurium LT2
SOURCE		Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
ORGANISM		1 (bases 1 to 20753)
REFERENCE		McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.
AUTHORS		Complete genome sequence of Salmonella enterica serovar Typhimurium LT2
TITLE		Nature 413 (6958), 852-856 (2001)
JOURNAL		21534948
MEDLINE		11677609
PUBMED		2 (bases 1 to 20753)
REFERENCE		The Salmonella typhimurium Genome Sequencing Project
AUTHORS		Direct Submission
CONSTRM		Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 444 Forest Park Boulevard, St. Louis, MO 63108, USA
TITLE		COMMENT Supported by NIH grant SU 01 AI43283
JOURNAL		Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/geneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES

source

```

1. .20753
  /organism="Salmonella typhimurium LT2"
  /mol_type="genomic DNA"
  /strain="LT2; SGSC 1412; ATCC 700720"
  /db_xref="ATCC:700720"
  /db_xref="taxon:99287"
  /note="LT2"
  complement(71..493)
  /gene="flgN"
  /note="synonym: STM1171"
  complement(71..493)
  /genes="flgN"
  /notes="believed to be export chaparone for FlgK and FlgP; flagella synthesis protein FLGN. (SW:FLGN_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20101.1"
  /db_xref="GI:16419688"
  /translation="MTRLSEILDQMTVLNDLKTVDAAEQQLSVGQINGSLQRLITE
  EKSLIADLYEQRLRQNAQRANDDIARWQAITKTOHLRLDNLQHNGLLEGG
  IERNQALVLPKPHQPTLYGADGQTSVSHRGKKISI"
  complement(498..803)
  /gene="flgM"
  /note="synonym: STM1172"
  complement(498..791)
  /genes="flgM"
  /notes="anti-sigma; negative regulator of flagellin
  synthesis (anti-sigma-28 factor). (SW:FLGM_SALTY); RflB
  protein"
  /codon_start=1
  /transl_table=11
  /product="anti-FlaA factor"
  /protein_id="AAL20102.1"
  /db_xref="GI:16419689"
  /translation="MSIDRTSLPKPVSTVOTRETSPTPVOKTRQKTSAAATSASVTL
  SDAQALMQQVSDINNRVEALKTAINRGLKMDIKIADSLIREAQSLQSK"
  complement(798..803)
  /gene="flgM"
  /notes="putative RBS for flgM; RegulonDB:STMS1H001438"
  complement(883..1556)
  /genes="flgA"
  /note="synonym: STM1173"
  complement(883..1542)
  /genes="flgA"
  /note="assembly of basal-body periplasmic P ring; flagella
  basal body P-ring formation protein FLGA precursor.
  (SW:FLGA_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20103.1"
  /db_xref="GI:16419690"
  /translation="MQTLKRGFAVALLPSPITMAQDINAQITTWFSQRLAGFSDEVV
  VTLRSPNLLPSCEQAFSGSAKLGNVNVVARCANEKRYLQNVQATGNVAVAA

```

```

PIARGGKLTANVTLKRGRLDQLPPTVLTDIRIQDAVSLRDLAPGQPVOLTMIRQAW
RVKAGQVQVIANGEGFVNAEQGMNAAVAQNRVMTSGQIVSGIVSDGNNILN
L"
  complement(1551..1556)
  /gene="flgA"
  /note="putative RBS for flgA; RegulonDB:STMS1H001439"
  1691..2115
  /gene="flgB"
  /note="synonym: STM1174"
  1691..1696
  /gene="flgB"
  /note="putative RBS for flgB; RegulonDB:STMS1H001440"
  1699..2115
  /gene="flgB"
  /note="cell-proximal portion of basal-body rod; flagellar
  basal-body rod protein FLGB (putative proximal rod
  protein). (SW:FLGB_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20104.1"
  /db_xref="GI:16419691"
  /translation="MLDRDLAALRFQOEALNLRQOEILANIANADTFGYQARDID
  FASELKKVMYRGREETGGVALTTTSSHIIPAQAVSSPAVDLLYRVPQPSLDGNTVDM
  DRERTQPADNSLKYQGLTVLGSQKGMNVLQGN"
  2105..2523
  /gene="flgC"
  /note="synonym: STM1175"
  2105..2110
  /gene="flgC"
  /note="putative RBS for flgC; RegulonDB:STMS1H001441"
  2113..2523
  /gene="flgC"
  /note="cell-proximal portion of basal-body rod; flagellar
  basal-body rod protein FLGC (putative proximal rod
  protein). (SW:FLGC_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20105.1"
  /db_xref="GI:16419692"
  /translation="MALLNIFDIAGSALAQSKELNVAAANLANADSVTPDGQPYRA
  KQVVFDAAPGQATGQVKVASVIESQAPKLYVEFGNPLADANGYKMPNVDVGM
  VNTMSARSQANIEVLNTVKSMLKLTLLGQ"
  2523..3233
  /gene="flgD"
  /note="synonym: STM1176"
  2523..2528
  /gene="flgD"
  /note="putative RBS for flgD; RegulonDB:STMS1H001442"
  2535..3233
  /gene="flgD"
  /note="initiation of hook assembly; basal-body rod
  modification protein FLGD. (SW:FLGD_SALTY)"
  /codon_start=1
  /transl_table=11
  /product="flagellar biosynthesis protein"
  /protein_id="AAL20106.1"
  /db_xref="GI:16419693"
  /translation="MSIAVNMDFNTGVTKTGSGMTGSNAADLQSSFTLLVAQL
  KNQDTPNPNLNETTOLAQISTVSGIEKLNTLGAISQIDNSQSLQATTLIGHYVM
  VPGTTLAGKABEGAVTSTPFGVLQOPADKVTAITDKDGRVVTLEIGELRAGV
  HTFTWDGKQDGTGTPNGSYNIALTSANGCTQLVAQLQALVQGVTKSGNGLDLG
  TYGTTLDEVRQI1"
  3248..4471
  /gene="flgE"
  /note="synonym: STM1177"
  3248..3253
  /gene="flgE"
  /note="putative RBS for flgE; RegulonDB:STMS1H001443"
  3260..4471
  /gene="flgE"
  /note="flagellar biosynthesis; flagellar hook protein

```


Enterobacteriaceae: Salmonella.
1 (bases 1 to 254050)
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
Stevens,K., Whitehead,S. and Barrall,B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
Nature 413 (6858), 848-852 (2001)
21534947
11677608
2 (bases 1 to 254050)
Parkhill,J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/s_typhi/).
Location/Qualifiers
1. .254050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
181. .2793
/gene="STY1078"
/note="synonym: pepN"
181. .2793
/gene="STY1078"
/note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
/codon_start=1
/product="aminopeptidase N"
/protein_id="CAD08183.1"
/db_xref="GI:16502232"
/db_xref="GOA:Q8Z770"
/db_xref="TrEMBL:Q8Z770"
/translation="MTQQPQAKYRHDYRAPDYQITDIETFDLDAEKTVTVAISOAVR
HGAPDLRLDGLTSLVHNDAPFWTAYKEEGALIIISDLPERFTLRVNEISPA
NVALEGLYSGDALCTQCAEGFRHTWYLDREDVLRFTTKIADKSKYPELLSNGN
RYAOLENGRHWQVQDEPPKPCVLPALVAGDFVLRDTFTTRSGRVVLELYVDRG
NDRAPWMTISLNSKMWDETRFGLDYLDITVIVAVDFPFGMANGENKGLNIFNSKYV
LARTDTATDKYDIDIERVIGHEYFHNWTVNTRCDWFQSLKEGLTVRQDEFSDDL
GSRVNRISNVTMRGLQPAEDASPMAPHRIDPKVIEMNNFYLLTVFKGAEVIRMIH
TLIGENFQKGMQLYFERHDGSAATCDDFVQMEDASNVYDLSHFRWYSQSQTPIVY
KDDNPETQYTLTISORTPATQDAEKOPHLPFAILEYDNEGNNVPLQKGHPVNA
VLNVTCAQRTFFDNVYFQVALLCERSAPVKLEYKWSQDLTFLMRHARNDPSRW
AQSSLATYIKLVNARHOGQPLSPVHVADAPRAVLIDDKIDPALAEILTPSANE
IAELFFYTAIQAUREALITRLAELADEFEIYNANHLDEYVYRDEGDKGRTLRN
ACRLFTFGETELANTLVSKQYRDANNMTDALAASAAVAQPCRDITLMQYDDKWH
OGLVMDKFIQSTSPAEVLETSLIKHESFMSNPNRISLIGAPAGSNPAAFH
AODGSGYQLVEMLTDLNRSNPQVASRLIEPLIRLKYRVDKQEKVRAALEQKLGLEN
LSGDLYEKIKTALA"
532. .1332
/gene="STY1078"
/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
family M1, score 245.10, E-value 1.9e-84"
1060. .1089
/gene="STY1078"
/note="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
3001. .4011
/gene="STY1079"

/note="synonym: pyrd"
3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli pyrd (PYRD_ECOLI); Fasta hit
to PYRD_ECOLI (336 aa), 93% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z7S9"
/db_xref="Swiss-Prot:Q8Z7S9"
/translation="MYIPFVRKALFQDPERAHEFTTQRRITGTPLEALVQKVPV
KPVTCMGLTFKNPLGAAGLDKDEICDALGAMFGSLEIGTVTPQPQNDKRLPFR
LVDAEGLINRMGNLGVNENVKAAHFDGLIGINIGKNKDTVPVNGKDDYLICME
KVYAGYVAINISSPNTGLRTLOYGDALDDLLTAIKKQNDLQVTHHKYVFAVKI
APDLCEELIQVADSLRHNDIGVIATNTTLDRLSVQGMKNCQOTGLSGRPLQKST
EITRRLSELKQGLPIGVGSDSVTAAREKIAAGATLVQIYSGFIKGPPLIKEIVT
HI"
3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODehase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238. .3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874. .3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174. .4719
/gene="STY1080"
4174. .4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="TrEMBL:Q8Z7S8"
/translation="MMRIKPDNNRWYVDEHDMVLDLNGMLFRSFRKMLTPDA
FCPTGCVDDAALVFSPEEKCRDPFELTKQRAELVNLVAIRVLPQMPKSHFVAH
GEMWFTGDAASVWLSDTAEQVNLVVEFGENAALCLLAQPGVVIAGRTMQLGDAIK
IMDRLKAPQVHCFSFSLEQAV"
complement (4716. .5825)
/gene="STY1081"
complement (4716. .5825)
/note="Orthologue of E. coli P75863; Fasta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q8Z7S7"
/db_xref="TrEMBL:Q8Z7S7"
/translation="MATLSRFLHPVSKMRGIGLTHALADISGLAFRIEMVTESDGT
FITARQPPQVVRFPFSLHGLHLPDGSALVRFVTFDPQAPTEVWGNHFFARVA
PTAINQWLSGFFRDVQLCWQFQLTRVRKRNAPLPGFADGYPLLTVNAGSLDLQO
RCPAGVQMEQFRNLVSGVAAWEEDSKVLRIGDVIYVVKPCSCRICFTVSPKQO
KHPGSEPLATLOAFRTAQDNGVDVFGONLIARNSGVIRVGVDEVEILATAPAKYGATT
LDQSVTEKHPDGSVTDWQGTFCGNNQVLLQLENQIGIRIPYSCRAGICGCCRIR
LLEGEVPLKSAAGDDGTILSCVFKTALRLN"
complement (4737. .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
5924. .8032

gene	RAINALCATPMDLMSILQNSQAGCLRAKLEII"	Score: 498.00	Matches: 144
	complement (5597. .6295)	Percent Similarity: 43.91%	Conservative: 76
	/gene="figh"	Best Local Similarity: 28.74%	Mismatches: 167
CDS	/locus tag="t1740"	Query Match: 19.26%	Indels: 114
	complement (5597. .6295)	DB: 1	Gaps: 14
	/gene="figh"	US-10-009-823A-1 (1-502) x AE016840 (1-301983)	
	/locus tag="t1740"		
	/notes="corresponds to STY1219 from Accession AL513382;		
	Salmonella typhi Crl8"		
	/codon start=1		
	/transl table=11		
	/product="flagellar L-ring protein precursor"		
	/protein_id="AA069364.1"		
	/db_xref="GI:29137803"		
	/translations="MKRYALHAYPMVWMLVAITLGCWIPAKPLVOGATTAQIPGPV		
	PVANGIFSAQINTGYQPLFEDRPRIIGTTLIVLQENVSASKSSANASRDGKT		
	SFGDTPVRYVLOGLFGNSRADMEAGNSFNGKGANASNTFSGLTIVTDQVLANGN		
	LHVVGKQAINQGTGFIRFSGVNPRTTISGNSVPSVQVADARIEYVGVNGYNEAQN		
	MGWLQRFPLNLSPM"		
	complement (6350. .7132)		
gene	/gene="figg"		
	/locus tag="t1741"		
CDS	complement (6350. .7132)		
	/gene="figg"		
	/locus tag="t1741"		
	/notes="distal rod protein; corresponds to STY1218 from		
	Accession AL513382; Salmonella typhi Crl8"		
	/codon start=1		
	/transl table=11		
	/product="flagellar basal-body rod protein FlgG"		
	/protein_id="AA069365.1"		
	/db_xref="GI:29137804"		
	/translations="MISSLWIAKTGLDQQTMTATNVLNANVSTNGFKRQRAVPEDL		
	LYTITQPAGSQSEQTLSGLQIGTGVSPVATERLHSGNLSTQNSKDVAKQGGF		
	FQWLPEDTSAYTRDGSFQVQDQVITAGGQVQVPAITIPANALISITIGRGVSVT		
	QOQCAPVQVQNLTFMNDTGESIGNLYETOSGAPNESPTPLNGALLLYGY		
	VETSNVAFELVNMIVQVRAVEINSKAVSTTDQMLQKLTL"		
	complement (7146. .7901)		
gene	/gene="figf"		
	/locus tag="t1742"		
CDS	complement (7146. .7901)		
	/gene="figf"		
	/locus tag="t1742"		
	/notes="corresponds to STY1217 from Accession AL513382;		
	Salmonella typhi Crl8"		
	/codon start=1		
	/transl table=11		
	/product="putative flagellar basal-body rod protein FlgF"		
	/protein_id="AA069366.1"		
	/db_xref="GI:29137805"		
	/translations="MDHAIYTANGAASQTLNQAVTASLANASTPGFRAQLNALRAV		
	PVDFGLATVLTASTPGADMTQGLDYTSRPLDVALQDDGLVYVQADGAEGYTRN		
	GNLQVPTGLTQGHPIVIGEGPIITVPEGSEITIAADGTISALNPGDPPTVAVPGR		
	LKLVKAEQNEVQSDGLPRLTAEAOERGAIVLAADPSIRINSGVLEGSNVKPVEMT		
	DMANARREPMQKVTITSVDENEGRANQLLSMS"		
	complement (7922. .9133)		
gene	/gene="fige"		
	/locus tag="t1743"		
CDS	complement (7922. .9133)		
	/gene="fige"		
	/locus tag="t1743"		
	/notes="corresponds to STY1216 from Accession AL513382;		
	Salmonella typhi Crl8"		
	/codon start=1		
	/transl table=11		
	/product="flagellar hook protein FlgE"		
	/protein_id="AA069367.1"		
	/db_xref="GI:29137806"		
	/translations="VSEFQAVSGLNAAATNLVDIGNNIANSATYGFKSGTAFADMEA		
	GSKVGVKRVAGITQDFTDGTINTGRGLDVALSQNGFFFLVDNSGVSFYSNGQFKL		

Alignment Scores:
Pred. No.:

1.22e-20 Length: 301983

```

Db      8335  -----ATGCGAGAGAAC----- 8324
QY      341  SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db      8323  -----ACC 8321
QY      361  SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgTyrSerGlnAsp 380
Db      8320  GGGCTAATAACATCGTCCCAACAT-----CAAAAC 8288
QY      381  GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly 400
Db      8287  GCGCTAAGCCCTGGCGACCTGCTGAGCTACCAAGATTAACAATGATGGACCGTGGTTCG 8228
QY      401  LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420
Db      8227  AACTACTCAACGACGACGAGCAGGCTGGGGCAGATGCTGCTGAATCTGCCAAC 8168
QY      421  GluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440
Db      8167  AACGAAGCTGCTGGCATCCAGGCGGATACGCTCGGGCGGACGACGCGCTCCGGGTA 8108
QY      441  ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
Db      8107  GCGCTGCTGGGACTGCGGCTTCCGGCACTTCGGTAACTGACGACGACGCGCGCTGGA 8048
QY      461  ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe 480
Db      8047  GCGCTCAACCTGATTTGAGTAAGACGCTGCTGTAATGCTGCTGCGCAGCGTAACAT 7988
QY      481  GlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500
Db      7987  CAGTCGATGCGCAGACCATCAAAACCCAGGACCATCCTCAATACGCTGTTAACCTG 7928
QY      501  Lys 501
Db      7927  CGC 7925

RESULT 27
AE016759 306358 bp DNA linear BCT 24-DEC-2002
LOCUS Escherichia coli CFT073 section 5 of 18 of the complete genome.
DEFINITION AE016759 AE014075
ACCESSION AE016759.1 GI:26107527
VERSION
KEYWORDS
SOURCE
ORGANISM Escherichia coli CFT073
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 306358)
Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P.,
Rasko, D.A., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J.,
Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C.,
Perna, N.T., Mobley, H.L.T., Donnenberg, M.S. and Blattner, F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
JOURNAL 12471157
PUBMED
2 (bases 1 to 306358)
Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P.,
Rasko, D.A., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J.,
Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C.,
Perna, N.T., Mobley, H.L.T., Donnenberg, M.S. and Blattner, F.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
JOURNAL
FEATURES
Location/Qualifiers
1.306358
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
repeat_region
gene
CDS

```

```

complement (83..3820)
/genes="iroC"
/locus_tag="c1253"
complement (83..3820)
/genes="iroC"
/locus_tag="c1253"
/notes="Residues 27 to 1231 of 1245 are 80.16 pct identical
to residues 1 to 1209 of 1218 from Genpept.129 :
>emb|CAD05883.1| (A1627276) putative ABC transporter
protein [Salmonella enterica subsp. enterica serovar
Typhim]
/codon_start=1
/transl_table=11
/product="ATP binding cassette (ABC) transporter homolog"
/protein_id="AAN79727.1"
/db_xref="GI:26107528"
/transl_table=11
/translation="MRDRSLAFNRNDPASGSYFTDGGGLMPANHTPTTQAQSWIVRLA
RVWERKSLVIVVASVSTLILAAITPLLTROAVNDALAGNAPRLPWLAAGLLIAF
FDIENRGVAGMLSLVQHTLRGRVDSIQKLDAGADALRTQVTSRTSDIQQ
VHLLQMCVPLAVFYIYTAGIAYLWSPMTLIVCVLCLAITALRARRVFAQT
GMASDQLANTEHIREVLQAGISVVKSCVAEMRETHWLDQSQROIIVRVRGAVISQAMP
GATMLALPVLQIVLICYGWSVMHGRIDLGITFVAFSLAMITGTRVLASLIVIAQ
RQASVVERFALIDTRSQMBEDGESINQVVGLENNMFYHHGDRHILSDISFSLR
AGTAVVAGASGKSTLMLLARFYDPCGKIWLNTSGRQNLRIEALRRVGI
VDFDAFCTVAENIAYGHPOATADDIRRAAAGASDFINALPKGFDLSLLTERGTN
LSGQRQRIALABALITAPDVLIDDTTSADVATEAINTALGRVADGHHMLLVIA
RRSTQLASRVVLDKGRVDTGTAELEAPCAFALMTGSDSDFLTSNNSNLWP
AEPATQDDVTIDGKGFVAMTRVFPENAVQQLAGKRVTSLLKPVAMFVIAALLI
ALDSAGVGLVILQHGIDSGAAGDMSTIGLCALALCLIVVWGCYSLSLQTVFAARA
AESQVSHVSRISFGHMLRLGHPWHRHADSRLTMTVDVDSLARFLONGLAGAATSLV
TFMAIATFWLDPILALTAISAPVAALATMYRRLSTPAPAQAARLEIGKVNSTIQE
KVSGLVQSHGQOEGELGARLRALSERFRATRVRAQKLYAVYFFLTFCTEASYAVAL
LVGASVAGENTAGVLAAPFLLLGOFYGPVQOLSGIVDAQOATASGKHIDELLATE
GTNLSGSSVLPVTGALHIDEVTFSPDSHEPALNKLTLTIPGVVAVVVGSGACKS
TLIKLAGLYPTHGNIRIGQMLDDASLITEYRQIGLVDDQVALFESSDIAENIRSR
PSATNEDVEIASQAGLYEMVCLPQGFPTVNNGGADLSAGORLIALARAQLANAH
ILLDEATSCLDRTSEERLMSLTVDVHAGKSHALIVAHRLTAAQRCDLIAVDKGLL
AEYGTHEQLISAGLYTRLMHDSVSTALHQRHNMKEETPG"
complement (3882..5045)
/genes="iroB"
/locus_tag="c1254"
complement (3882..5045)
/genes="iroB"
/locus_tag="c1254"
/notes="Escherichia coli O157:H7 ortholog: z1190"
/codon_start=1
/transl_table=11
/product="putative glucosyltransferase"
/protein_id="AAN79728.1"
/db_xref="GI:26107529"
/transl_table=11
/translation="MRRPLDQRABDFLMRLIFVGPPLYGLLYPVLSLAQAFVNG
HEVLASGGQFAQAAEAGLVDFDAAPGLDSEAGYRHEAQRKSNIGTQGNFSFES
EENADHIVEFAGHWRPLIIYPLGVIGPLIAAKYDIPVVMQTVFGHTPWHIRGVTR
SLTDAYRHNVGATPRDMAWDVTPPSMSTILENDGPEIIPMOYVPYGGAVWPBWER
RPRKRLVSLGTVKQVNDGLDLIAWMSASVDAEILILHISANASDLRSPLSNVR
LVDMIPMGVFLNGADGFIHHGAGNTILAHAGIPQIVFGGADRPVNAVRAVGGG
IIFDVLGSLNMFANFLNNRSLRKASAEVAEAAQPCPGEVAKSLITMTVQKG"
5535..5834
/locus_tag="c1255"
5535..5834
/locus_tag="c1255"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79729.1"
/db_xref="GI:26107530"
/translation="MKPTLLQTVENICQLIFVLFTGTGLIDVEALVSCSSVAKNQ
RTPASSVSSSVASVSRTAADIMLLSNPLFAPGVSPFGCSLAASGIHVTILE"
5867..6577
/insertion_seq="IS1541A-like"
5961..6470
/locus_tag="c1256"
5961..6470

```

```

/locus tag="c1256"
/notes="Residues 18 to 169 are 94.07 pct identical
to residues 1 to 152 of 152 from SwissProt.40 :
>sp|Q57334|T200_SALTU Transposase for insertion sequence
element 18200"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79730.1"
/db xref="GI:26107531"
/transl_table=11
/translation="MRSQCKCQTRNQKQVPMGNEKSLAHTRWCKYHIVPAPYRRQ
VFYREKRAIGCILRLCEWKSRIIEACECCADHIMLVEIPPMVSFGFMGLKGS
SLMPYQFQGLKFKYRNREFWCRGYVDVTGKNTAKIQDYIKHQEEDKMGQLSIPY
PGSPFFTRK"
/complement(6578..6940)
/notes="incomplete"
/insertion seq="IS4"
/complement(6598..6960)
/locus tag="c1257"
/complement(6598..6960)
/locus tag="c1257"
/function="IS, phage, Tn"
/notes="Residues 8 to 120 of 120 are 98.23 pct identical to
residues 316 to 428 of 442 from MG1655 : b4278"
/codon_start=1
/transl_table=11
/product="Putative conserved protein"
/protein_id="AAN79731.1"
/db xref="GI:26107532"
/transl_table=11
/translation="MPROPDQGYREIKOTMOLSRITLRSKPKBELVBOELWGLVLA
NYL
VYQMLKMAEHLKGYMNPOLSFESCGMYMELMTIQGASPRGIPELMRDLASMGQV
KLPTRETAPEPRVVKERP"
7651..8013
/locus tag="c1258"
7651..8013
/locus tag="c1258"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79732.1"
/db xref="GI:26107533"
/transl_table=11
/translation="MLGACFPYSWGTSFYXGFPWMGAELVIMMWPEACFLLQSSLYSL
LDQGVLPFSFGLNKQYEGFEFYQIFSGNIVASCSSCGSVSSNAHYARCTDFWRH
YSLNVAACOLGRQACLS"
7835..8269
/locus tag="c1259"
7835..8269
/locus tag="c1259"
/notes="Residues 1 to 139 of 144 are 50.71 pct identical to
residues 104 to 242 of 258 from GenPept.129 :
>gb|AAK18487.1|AF348706_176 (AF348706) orf, hypothetical
[Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79733.1"
/db xref="GI:26107534"
/transl_table=11
/translation="MKDLSFIRPFLLAILLLVPPALEFFVIHRRITPGALTLCVITAL
IITLLASGLRYSVKCISELSFIRBRIECMWMAGPMIYFTFSLIAGWKYQFKEY
QTHYENARYIVLARDYDGLRLVLSQYSRSRKFVIINGHRL"
8477..8553
/locus tag="c5534"
8477..8553
/locus tag="c5534"
/product="tRNA-Arg"
/function="tRNA-Arg"
/complement(8832..9290)
/locus tag="c1260"
/complement(8832..9290)
/function="putative enzyme; Transposases Insertion
Sequence Associated"
/notes="Escherichia coli O157:H7 ortholog: z4324"

```

```

/codon_start=1
/transl_table=11
/product="Putative Transposase"
/protein_id="AAN79734.1"
/db xref="GI:26107535"
/translation="MASAADARRIVSHYERRWLEEVHKWAKSGGTCVESLRMQTRDN
LERMVVIAFAVLMGLRQEG-SEETONDSCKILPTTEWKLWKLKGLQSPQP
TLKWACLKGRWHDSKRTGPRGWVWMDGNFRLQDMVEGYPVMSKLDQET"
/complement(9302..9667)
/locus tag="c1261"
/complement(9302..9667)
/locus tag="c1261"
/notes="Residues 1 to 35 of 121 are 97.14 pct identical to
residues 59 to 93 of 316 from GenPept.129 :
>gb|AAK18437.1|AF348706_126 (AF348706) putative
transposase [Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79735.1"
/db xref="GI:26107536"
/translation="MABRLGEIQKRVITVCDREADIWHYLYKVSHGQREVFVTAQNS
US-10-009-823A-1 (1-502) x AE016759 (1-306358)
Alignment Scores:
Pred. No.: 1,248-20 Length: 306358
Score: 498.00 Matches: 143
Percent Similarity: 43.57% Conservative: 74
Best Local Similarity: 28.71% Mismatches: 171
Query Match: 19.26% Indels: 110
DB: 1 Gaps: 15
QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db 73424 AGCGGATTAAACGCTGCCGCCACCACCTCGATGTTATTGGCAACAATATCGCACTCC 73483
QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 73484 GCCACCTACGCTTTAAATCAGGCACGCGCTCTTTCCGATATGTTTGGC----- 73534
QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 73535 -----GGTTGG-----AACTGGGACTCGGGGTAAAGTTGCT 73567
QY 70 SerValargThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 73568 GGTATCACTCAGGACTTTACCGATGCGCAGCACCACACCGCGGGCGTGTCTGGAGCTT 73627
QY 90 AlaIleGlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr 108
Db 73628 GCTATCAGCCAGAACGGTTTTTTCGCTGTGTAGACAGCAACGGTTCGGTATTCTACAGC 73687
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 73688 CGTAACGGACAATTAAGCTGGATGGAACCGTAACCTGTGATATGCAAGGTTTACAG 73747
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
Db 73748 CTGACGGGTACCGGCAACCGGTACCGCCGCGACTATTTCGCAACGCGGCGGAATCCGACC 73807
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 73808 AATATTTTCGATCCCGAATACCTCGATGCGCA-----CGGAAACTACCAACGCGG 73858
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 73859 TCGATGAGATCAACCTG---AATTCAGTATGCTGCTTCCCTCTGTT----- 73903
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyr 208
Db 73904 -----AACGCATTTGATCCCAATGCGGATGCTAT 73936

```


QY	109	ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr	128
Db	7431	CGTAACGGCAAAATTAAGCTGGATGAAATCGTAACCTGCTGAATATGCAAGGTTACAG	7490
QY	129	LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle	148
Db	7491	CTACGGGTGTACCCGGCAACCGGTACCCGCCGACTATTTCGACGAGGGCGAATCCGACT	7550
QY	149	GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu	168
Db	7551	AACATTTCGATCCGGAATACCTGATGGCA-----CGGAAAACCTACCAACGCGCG	7601
QY	169	AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr	188
Db	7602	TCGATGCAGATCAACCTG---AATTCAGCGATCCGCTCCCTCTGTT-----	7646
QY	189	PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyr	208
Db	7647	-----AACGCAATTTGATCCGCAATGCGGATAGCTAT	7679
QY	209	SerTyrAlaGlnProMetArgValTyrAspGlnGlnLysAsnSerHisAspIleThrVal	228
Db	7680	AAAAAAGGTTCCGGTGACTGTTTCGACAGTCAGGTAAATGCTCATGACATGAGCGTC	7739
QY	229	TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet	248
Db	7740	TACTTT-----GTGAAGACCGGGATAATACTGGCAGGCTCTAC-----	7778
QY	249	AsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSer	268
Db	7779	-----ACCCAGGATAGCATGATCCAACAGGTACAGCGAGCGCTGCAATG-----	7823
QY	269	GlyThrMetThrPheSerSerAsnGlyGluLeu---LysAsnMetThrAlaPheThrPro	287
Db	7824	---AAGCTGGTGTTTAATGCCATGGCGTCTGCACCTCAAATCCACAGAGAATATTACC	7880
QY	288	ThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeu	307
Db	7881	ACCGGCGCA-----ATTAAACGGCGCAACCCGCC-----	7910
QY	308	ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly	327
Db	7911	---ACGTTTAGCTGAGCTTCTCACTCC-----	7937
QY	328	IleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThr	347
Db	7938	-----ATGCAGCAAAAT-----	7949
QY	348	AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla	367
Db	7950	-----ACCGCGCTAACACATTTGGCA	7973
QY	368	ArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeu	387
Db	7974	ACC-----ACCGAATGGCTACAAACCGGCGCATCTG	8006
QY	388	ValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnVal	407
Db	8007	GTGAGTTATCAATCAATGATGAGTACGGTTGTCGGCAACTATTCCACGACAAACC	8066
QY	408	ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu	427
Db	8067	CAACTGCTGGGCGAGATTGTACTGGCGAACTTTGCCAAACACGAAGGCTGGGATCCGAA	8126
QY	428	GlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly	447
Db	8127	GGCGAACAGTCTGGTCTGCGAGCAATCTTCTGGCGTGGCGCTGTGGGGACAGCGCGG	8186
QY	448	ThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSer	467
Db	8187	ACGGGCAACTTTGGCACCCCTGACCAACCGGTGCGTGGAGCGTCCACGCTCATCTCAGT	8246

Qy	468	ArgGluMetValAsnVetIlelleIleGlnArgGlyPheGlnMetAsnSerLysSerVal	487
Db	8247	AAAGAACTGGTCAATATGATCGTTGGCCAGCGTAACTATCAGTCTAACGCCAGACCATC	8306
Qy	488	ThrThraAlaaspThrMetLeuGlnLysAlaLeuGluLeuLys	501
Db	8307	AAACCCAGACAGATCTCTCAACACGCTGGTTAACTTACGC	8348
RESULT	29		
AP002555			
LOCUS			
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 6/20.		
ACCESSION	AP002555	BAC00007	
VERSION	AP002555.1	GI:13360886	
KEYWORDS			
SOURCE	Escherichia coli O157:H7		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1		
AUTHORS	Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.		
TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak		
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)		
MEDLINE	20198780		
PUBMED	10734605		
REFERENCE	2		
AUTHORS	Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shingawa, H. and Hayashi, T.		
TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MGL655		
JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)		
MEDLINE	20557356		
PUBMED	11108008		
REFERENCE	3		
AUTHORS	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.		
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak		
JOURNAL	Gene 258 (1-2), 127-139 (2000)		
MEDLINE	20564182		
PUBMED	11111050		
REFERENCE	4		
AUTHORS	Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.		
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12		
JOURNAL	DNA Res. 8 (1), 11-22 (2001)		
MEDLINE	21156231		
PUBMED	11258796		
REFERENCE	5 (bases 1 to 222605)		
AUTHORS	Hattori, M., Ishii, K. and Shiba, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2000) Masahira Hattori, Kitasato Institute for Life Sciences, Kitasato University; Kitasato 1-15-1, Sagamihara, Kanagawa 228-8555, Japan		
	(E-mail:hattori@genome.ls.kitasato-u.ac.jp, URL:http://genome.ls.kitasato-u.ac.jp/, tel:81-42-778-8194, Fax:81-42-778-8193)		
COMMENT	This work was done in collaboration with Tetsuya Hayashi, Makoto genome project		

Ohnishi, Keisuke Nakayama (Miyazaki Medical College), Kozo Makino, Ken Kurokawa, Katsushi Yokoyama, Masashi Tanaka, Takeshi Honda, Teruo Yasunaga, Hideo Shinagawa (Osaka University), Takahiro Murata (Shinshu University), Chang-Gyun Han, Eiichi Ohtsubo, Toru Tobe, Chihiro Sasakawa (University of Tokyo), Hideto Takami (Japan Marine Science and Technology Center), Naotake Ogasawara (Nara Institute of Science and Technology), Satoru Kuhara (Kyushu University), and supported by the Research for the Future Program of the Japan Society for the Promotion of Science.

FEATURES

source

1. 222605
Location/Qualifiers
/organism="Escherichia coli O157:H7"
/mol_type="genomic DNA"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
74. .607
/gene="ECs1423"
74. .607

/gene="ECs1423"
/note="similar to B1045 ECOLI gi|1787283 percent identity 100 in 177 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative polyprotein"
/protein_id="BAB34846.1"
/db_xref="GI:13360887"
/translation="MKTRHVGQDITKLAVGVVNAAPSLMGGGVGATHRAAGP
ALLDACLVKQQDCPTGHAVITLAGDLPKAVHTVGFVWEGSEQNDQLDAYL
NSRLNAANSVTSVAFAISTGVGYPRAAAEIAVKTVSFEITRHALPEQYIFVCYD
EENAHYERLLTQGDSE"
549. .2030
/gene="ECs1424"
549. .2030

gene

CDS

/gene="ECs1424"
/note="similar to YMC_ECOLI gi|1787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative synthase"
/protein_id="BAB34847.1"
/db_xref="GI:13360888"
/translation="MMKKTPTKDSLPNKEMNDLPRLASVLPICSPQPGCGLPPL
EKSLDAFAARYLAEMSEHTLDVQYIWDMSGRLLFSAALAAKRGVVRLLDDN
NPLGDIALLDSHPRIEVLNPFSLRLLPLGYITDFSLNRNMKNKSTVGVV
TLVGRNIGDAYFCAGEEPLFSLDVAIGPVVEDVADDFARYWCKSVPLQQVLVD
PEGEMADRIELPASWINDAMTHRYLRKMESPPIHNLVDGTLPLIWAQTRLLSDDPK
GEKAKHSLPDLFDINGSPERDIISSVFPVTRAGVOLLRWKGVKIAIITN
SLANDVAVHAGYAKRWKKLLRYGVELHELKRETSSTLHDEGTGNSGSLHAKT
FSIDGKTVFGSFNPRSTLLTEMGFVIESETLAQLDKRPIQSCYDAQAQLRLDR
WGRINWDRAKKEIIVLKEPATSFWKRVNVLASILPVEWLL"
complement(2038. .3195)
/gene="ECs1425"
complement(2038. .3195)
/gene="ECs1425"
/note="similar to MDOC_ECOLI gi|3025226sp|P75920 percent identity 99 in 385 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="glucans biosynthesis protein"
/protein_id="BAB34848.1"
/db_xref="GI:13360889"

gene

CDS

/translation="MNPVPAQRYFLDSTRANMLLIGIPHISLIYSSTWHVNGAEP
SLMLTFNDHFMQVFFVIGSYFMYLFLRYPLKWKWKVVRVGPIMPLTPLL
TLQFIMLVYVKGASWPGSLGYKYNLAWELISHLWFLVLVMMTLCVIMFKRI
RNNLSDNKTVMKSVFELCLGIGYAVIRRTFIVYPIILSGNFVIMQTL
FYLPFFTLGALATFPHLKALFTTPRGCTLAALAFVAYLLNRYSGDAMWYTES
VTWVLGLVMNVVSGFRLNFOQARTVFNASLIYLVHHPJTLFPFGAYITPHI
TSNWLGLFGLGILFVWGTAITLYEHLRILPLKFLFSKGVKRENDKAPAR"
3588. .5123

gene

/gene="ECs1426"
3588. .5123
/gene="ECs1426"
/note="similar to MDOC_ECOLI gi|1787286 percent identity 99 in 511 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="periplasmic glucans biosynthesis protein"
/protein_id="BAB34849.1"
/db_xref="GI:13360890"
/translation="MMKORWLSAAVMJLTSSSWAFSIDVAKQASLAGKYEAPK
SNLPSVFRDMKYADYQIQFNHDKAYNNLUKTFKLEFYHQGYFDTFVKINEVTATA
VKRIKSPDYPTFGDVQDKDVTGDFAGFKVLYIPINSKDKNDIIVMLGASYFRVI
GAGVYGLSARGLAIDTALPGEFPRFKEWIERPKPTDKRLTIYALLDSPRATGAY
KFVMPGRDVTVDVQSKIYLRDKVGLGVAPLTSMFLGPNOPSPANNYRPELHDNG
LSIHAGGENIWRPLNPKHLAVSSPSMENPQGLLQRCDFSRFEDLDRLDRLPS
AWPTPKGEMGKSGVELVEIFTNEDNINIVAYTWPOLPEPKEMAFKVTITPSRDED
KLHAPDNAAVQQRRTSGDVQSNLRQPQGTIAFVVDFTGAEMKKLPEDTPTVATQTS
IGDNGEIVESTRYNPVTGKRLVMRVKVDKAKKTTEMAALVNAQDTLSTWYQLP
ANE"
5086. .7659
/gene="ECs1427"
5086. .7659
/gene="ECs1427"
/note="similar to membrane glycosyltransferase MDOH_ECOLI
gi|1787287 percent identity 100 in 847 aa but differs at
N-ter (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="membrane glycosyltransferase"
/protein_id="BAB34850.1"
/db_xref="GI:13360891"
/translation="MKPGATSYLPNMKTTIYIDAMPIAASEKALPKTDIRAVHQALD
AEHTWAREDDSPQGSVKARLEQAPDSIADQLIKDDEGRDOLKAMPEAKRSMPFD
PWRTVGRFWRDLRGDRDTPRYLARTKEEQQSEOKWRTVGTIRVILLITLAQTV
VATYMKITILPYQGHALLNPMQVGDLYWSEMQLLPYMLQGTGILLFVFLFCWVSAG
FWTALMGFLQLIGRDKISASTVGDPEPLNPHETALINPICNEDVNRVFAEGLATM
ESVATGNKAKHFDVILDSINPDICVAEGKAWELIAEVGGGQIFRRRRRVKRR
SGNIDDFCRWGSQSYMYVLDADSVMTGCLCLVRLMEANPAGIIQSSPKASGMD
TLVACQOFATRVYGPITAGLFWQGESHYGHNAIIRKPFPIEHALAPLPGES
FAGSILSHDFVEALMRAGWVMIAYDLPGSEELPNLLDELKDRDCHGNLNNF
RLTVKGMHPVRAVFLITGVMSYLSAPLWFMFLALSTALQVHALTEPOYFLOPRLF
PWPQWPEALIALFASTWLLFLPKLLSILLNCKTGKEYGGFWVTLISLLEVLFS
VLLAPVRLMFTVFNVSAPLGEVWVNSPQRDDSTSGEAFKRGHQLLGLVWVG
MAWDLFLFWLAPVFSLLSPSVISSRATVGLTKRWKLLFLPEEYSPQVLLVD
TDRELMNRQRSDDDGFMAFVNPFSFALATAMATARHASKVLEIARDHRBQALNE
TPEKLNDRILVLLSDPVTMARLHFRVNSPERSYSSWVSYEGIKLNLPLALRKPDAAS
Q"
7832. .8059
/gene="ECs1428"
7832. .8059
/gene="ECs1428"
/note="identical to YCEK_ECOLI gi|1787288 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB34851.1"
/db_xref="GI:13360892"
/translation="MRLIVVSIWVTLSSGCGSIISRTIPQGGHGNQYYPGVQWMDVRDS
AWRVYTLIDLPFSLVFDLTLLLPIDIHGPEY"

Alignment Scores:

Pred. No.:	1.8e-20	222605
Score:	493.00	142
Percent Similarity:	44.13%	76
Best Local Similarity:	28.74%	174
Query Match:	19.06%	102
DB:	1	15

QY	10	ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla	29
Db	26851	AGCGGATTAACACGCTGCCGCCAACCACTCGATGTTATTGGCAACAATATCGCAACTCC	26910
QY	30	AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu	49
Db	26911	GCCACCTACGGCTTTAAATCAGCAGCGGCTCTTTTGGCGATATGTTGCC	26961
QY	50	AlaIleGlySerThrClySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly	69
Db	26962	-----GGTTCCG-----AAAGTGGGACTCGGGTAAAGTTGCC	26994
QY	70	SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu	89
Db	26995	GGTATCACTCAGGACTTTACCGATGCGACGACCAACACCGGCGGTCTCGACGTT	27054
QY	90	AlaIleGlyClyLysGlyPhePheGln--ValThrLeuGluAspLysValHisTyrThr	108
Db	27055	GCTATCAGCCAGCAACGGTTTTTTCGTCTGCTAGATAGCAACGGTTCGGTGTCTTACAGC	27114
QY	109	ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr	128
Db	27115	CGTAACGGACAAATTTAAGCTGGATGAAATCGCTAACTGGTGAATATCAAGGTTTACAG	27174
QY	129	LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle	148
Db	27175	CTACGGGTTACCGGCAACCGGTACGCGCGCACTATTTCACGAAGGGCGAATCCGACT	27234
QY	149	GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu	168
Db	27235	AACATTTCGATCCGAATACCTGATGGCA-----GCGAAACTACCAACACGCGC	27285
QY	169	AsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr	188
Db	27286	TCGATCGCACTCAACCTG---AATTCACGGATCCGCTCCCTCTGTT-----	27330
QY	189	PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyr	208
Db	27331	-----AACGCATTGTATGCCCAATATCGGATAGCTAT	27363
QY	209	SerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrVal	228
Db	27364	AACAAAAAGGTTCCGTGACTGTTTCGACAGTCAGGGTAATGCTCATGATGAGCTC	27423
QY	229	TyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMet	248
Db	27424	TACTTT-----GTGAAGACCGGGGATAAATCTGCGAGCTCTAC-----	27462
QY	249	AsnProSerGlnAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSer	268
Db	27463	-----ACCCAGATGACGTGATCCACAGGTACAGCGACCGCTCGCAATG-----	27507
QY	269	GlyThrMetThrPheSerSerAsnGlyGluLeu--LysAsnMetThrAlaPheThrPro	287
Db	27508	---AAGCTGGTGTTTAATGCCATGCGTCTTGACCTCAATCCACAGAGAATATTACC	27564
QY	288	ThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeu	307
Db	27565	ACCGGCGCA-----ATTACGGCGCGAATCCCGCC-----	27594
QY	308	ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly	327
Db	27595	---ACGTTTAGCTGAGTTCCTCAACTCC-----	27621
QY	328	IleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThr	347
Db	27622	-----ATGCAGCGAAT-----	27633
QY	348	AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla	367
Db	27634	-----ACCGGCGCTTAAACAACATTTGTGCA-----	27657

RESULT 30
U00096_11
WPCOMMENT

U00096_39 3900001 4010000
 U00096_40 4000001 4110000
 U00096_41 4100001 4210000
 U00096_42 4200001 4310000
 U00096_43 4300001 4410000
 U00096_44 4400001 4510000
 U00096_45 4500001 4610000
 U00096_46 4600001 4639675
 Continuation (12 of 47) of U00096 from base 1100001 (U00096 Escherichia coli K-12 MG1655)

Alignment Scores:
 Pred. No.: 9,78e-21 Length: 110000
 Score: 491.50 Matches: 143
 Percent Similarity: 43.95% Conservative: 75
 Best Local Similarity: 28.83% Mismatches: 173
 Query Match: 19.01% Indels: 105
 DB: 1 Gaps: 15

US-10-009-823A-1 (1-502) x U00096_11 (1-110000)

Qy 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
 Db 31818 AGCGGATTAACCGTGGCGCACCACTCGATGTTATTGGCAACAATATCGCAACTCC 31877
 Qy 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
 Db 31878 GCCACCTACGCTTTAAATCAGGACCGCTCTTTTCCGATATGTTGCC----- 31928
 Qy 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
 Db 31929 -----GGTTCG-----AACTGGAGCTGGGGGTAAAGTTGCC 31961
 Qy 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
 Db 31962 GGTATCTACTCAGGACTTTTACCGATGGCGACGACCAACCGCGGAGGTCTCGGACGTT 32021
 Qy 90 AlaIleGlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr 108
 Db 32022 GCTATCAGCAGACGCTTTTTCGTCGTGTAGACAGCAACGCTTCGTTGTTCTACAGC 32081
 Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 Db 32082 CGTAACGGCAATTTAAGCTGGATGAAACCGTAACCTGTGTAATGCAAGGTTTACAG 32141
 Qy 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlyLysGluThrLeuGluProIle 148
 Db 32142 CTGACGGTTTACCGGCAACCGGTACGCCGCCCACTATTACGCAAGGGCGAATCCGACC 32201
 Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
 Db 32202 AATATTCGATCCGAATACCTCGATGGCA-----CGCAAAACTACCAACCGGCA 32252
 Qy 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 Db 32253 TCGATGAGATCAACCTG---AATTCAGTGATCCGCTTCCTACT----- 32294
 Qy 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProPheLysSerThrSerAsn--- 207
 Db 32295 -----GTTACGGCAATTCAGCGCCGCAATGCG 32321
 Qy 208 -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnLysAsnSerHisAsp 225
 Db 32322 GATAGCTATAACAAAAGGTTTCGATGCTTTTCGACAGTCAAGGTAATGCTCATGAC 32381
 Qy 226 IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu 245
 Db 32382 ATGAGCTCTACTTT-----GTGAAGACCGGGGATAATAACTGGCAG----- 32423
 Qy 246 ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu 265
 Db 32424 GTCTACACCGAGTACAGTGTATCCAAACAGCATTCGGAAGACA----- 32468
 Qy 266 LeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe 285

Db 32469 -----GGCAACAACACTGGAATTTAATGCTAATGGCACATTA-----GTGATGGTGGC 32516
 Qy 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
 Db 32517 ATGGCGAATATATCGCAACCGCGCAATTAACGGTGCAGAACCCGCC----- 32564
 Qy 306 GlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAsp 325
 Db 32565 -----ACGTTTAGTCTGAGCTTCTCAACTCC----- 32591
 Qy 326 PheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIle 345
 Db 32592 -----ATGCAGCAAAAT----- 32603
 Qy 346 GlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSer 365
 Db 32604 -----ACCGGCGCTAACCAATATT 32621
 Qy 366 ThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGly 385
 Db 32622 GTGGCAACC-----ACCCAGAACGGCTACAAACCCGGGC 32654
 Qy 386 AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSer 405
 Db 32655 GATCTGGTGGTATCAATCAATGATGACGTCAGCTGTGCGCAACTATTCACACGAA 32714
 Qy 406 GlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 425
 Db 32715 CAAACCAACTGCTGGGCGAGATGTACTGGCGAACTTTGCCAACAGAGGTCTCGCA 32774
 Qy 426 ArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu 445
 Db 32775 TCCGAAGCGCAACGCTGTGCTGCGACCAATCTTCTGGCGGCGCTGTGGGACA 32834
 Qy 446 ProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAsp 465
 Db 32835 GCCGGGCGGAAACTTTGGCACCTGCACCAACCGTGGCGCTGGAAGCGTCCACGTCAT 32894
 Qy 466 MetSerArgGluMetValAsnMetIleIleGlnArgPheGlnPheGlnMetAsnSerLys 485
 Db 32895 CTCAGTAAGAACTGGTCAATATGATCGTCCCGCGTAACATCATCTTAACGCCCCAG 32954
 Qy 486 SerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 Db 32955 ACCATCAAAACCCAGGACAGATCCTCAACACGCTGGTTAACTTACGC 33002

RESULT 31
 AE001509/c
 LOCUS
 DEFINITION Helicobacter pylori, strain J99 section 70 of 132 of the complete genome.
 ACCESSION AE001509
 VERSION AE001509.1 GI:4155350
 KEYWORDS
 SOURCE Helicobacter pylori J99
 ORGANISM Helicobacter pylori J99
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.
 1 (bases 1 to 19549)
 Aln,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C., Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G., Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C., Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E., Vovis,G.F. and Trust,T.J.
 Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
 Nature 397 (6715), 176-180 (1999)
 JOURNAL
 MEDLINE
 PUBMED
 9923682
 REFERENCE 2 (bases 1 to 19549)
 AUTHORS King,B.L., Alm,R.A. and Trust,T.J.
 TITLE Direct Submission

JOURNAL	Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA
COMMENT	Address all correspondence to: hp@arch.us.astra.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Jo-See L. Ling, Donald T. Moir, Douglas R. Smith, Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: http://www.astra-boston.com/hpylori).
FEATURES	Location/Qualifiers
source	1. 19549 /organism="Helicobacter pylori J99" /mol_type="genomic DNA" /strain="J99" /db_xref="taxon:85963" complement(309..932) /genes="hds_2" /notes="hds_2" /notes="synonym: jhp0785" complement(309..932) /genes="hds_2" /notes="similar to H. pylori 26695 gene HP0848 and HP0849" /codon_start=1 /transl_table=11 /product="TYPE I RESTRICTION ENZYME (SPECIFICITY SUBUNIT)" /protein_id="AAD06377.1" /db_xref="GI:4155365" /translation="MNKIRLLQTLAPKGVGRKKEVCESNNKTLKISEVSEVKNK GMPVINSGRELYGYHDFNNGENITIASRGEYAGFINFNEKFPAGLCYPPYKVD TNELTKFYLYKTNQETIMENLVFRGSIIPALNKADITETLPIPPLEIQOEIVTIL DQFSALTDLQAGIPAEIKARKQYEEYREKLLAFKPLTPLNNKELA" complement(943..2529) /gene="hdsM_2" /note="synonym: jhp0786" complement(943..2529) /genes="hdsM_2" /notes="similar to H. pylori 26695 gene HP0850" /codon_start=1 /transl_table=11 /product="TYPE I RESTRICTION ENZYME (MODIFICATION SUBUNIT)" /protein_id="AAD06378.1" /db_xref="GI:4155366" /translation="MENTNKASAKLNHNSLARNELHNTIWKVANELEGSDGWDFFQ YVLGILFYRISNMNHNVEYRKIDPNFDVASIDDEAEIVRKSTIEKEGFFIPPS ALFCNVLKAPNNEIDLNVLTONTFNEIEKSLGTPSENVKGLFADLDVNSKLGSSH QNRVEKLTIQALQGMQLDYOQSGIDVFGDAYEYLMAMVSNACKSGEYETPOEV SELAKITLHNQENINKVDPCCGSGSLLLQFSKVLGDKXNLKGFQSGEINLTNNLC RINFLHDINSKFHIALGDTLLDPKHEDDEPFDALVSNPPYKRWIGDNPNLLWDE RFNKAGALAPKNAADLAFTMMLSYLSNGAAIIVFPFGVLYRSAGEKKIRYLVKEN FIDCVIALPENLFFGTNIATCILVKKNKDDTTLFIDASKFELKEGKNLKAHRE KILQTVIERKAIKHFSAIANIEKIQENDYNLSVNRVVEQDTEKIIDIKALNGEISQI VKQSVLRNSLESIIKELEI" 2708..3394 /gene="jhp0787" 2708..3394 /gene="jhp0787" /notes="similar to H. pylori 26695 gene HP0851" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06379.1" /db_xref="GI:4155367" /translation="MVFDRTTSVREKKAALGIVGIVFFILFGIVISGASQKQWVQ QLDLFDILIRNPAPIQGSWLSFVFWFAQSCLTPIALLIGLWFGFQRIALGV WFFPSILGGEFTLKLVARPRTVNGELVFAGHGFSPSGHALASALFYGSLLALL CYSNANNRKITGAILLFWILMAYDRVYLVGHVPSDVLGGFLGIAWSCCSLALYL
gene	GFLKRPYKAA complement(3437..4378) /gene="jhp0788" complement(3437..4378) /gene="jhp0788" /note="similar to H. pylori 26695 gene HP0852" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06380.1" /db_xref="GI:4155368" /translation="MKPDIEIIQSLEALEKEPKVTEIYDKAKELPEKEIGESMPFY GGETPHOSVSAAYITALNKGEEPLFKVQEKPVLIALLKGAKEPVNLIEKGVSSVKI VHKIMERDLHPITTYMAYINENLKYTKITIEHSESSKSPKMDRWLYPDMVGVRFL HAEUSLENLAFSKKFTDLPILKLVSEFKEKIEIVSHNCRECYFOAISNSVAENGLYVG RHIDTNPQLMDLKLRLHSGFVGIDLTREDKSAILLNNAKHEKIDITYVALELSAK NEKPSGLKSVVDYDNNHPRYKDFEIDKKEELYP" complement(4375..5976) /gene="jhp0789" complement(4375..5976) /gene="jhp0789" /note="similar to H. pylori 26695 gene HP0853" /codon_start=1 /transl_table=11 /product="putative abc transporter, ATP-binding protein" /protein_id="AAD06381.1" /db_xref="GI:4155369" /translation="MLQTIINLTQRYATKLFENNVNKLKDKRYGLIGANGAGKSTFL KIKSIDSCESEVITITSGMRMVLGQDQAFEDLSLKDVLIGNRLYDAIKEREL YTEGLSDSKVGNARLGELETICVEEDPMYSEVIEKLEDLGIIPSSKNDLMKTLPS SDKFILDLAQLVLPKPDILLDEPTNNLDLNAIEMLENNKRHEGTWVISHDRFLN AVCETHILDQPHSVREPSGNVDWYIASTLIAKQEAERNKLKEKELEKFTARFSA NASAKATSQKOLDKLDIGSLAISRRSPSIIFKPKRTIGNEALCEMISKSYDDQ IVLNQSLKMPKDKIALIGPNVGKSTLCKILVEELKPDGTGVVKGATVLKGYEPPN VSEISRETYIQWLFNFKKISAEVNRNALGRMLFNGEQEKCVNALSQGEKHEMVL LKVPKSGASIDFEPSTNHLDELAITIALGEALFKPDGALLICVSHDRDLIDAYANRIIE LVPSPKASIDFKGSYEYLASKK" 6178..7155 /gene="guaC" /note="synonym: jhp0790" 6178..7155 /gene="guaC" /note="similar to H. pylori 26695 gene HP0854" /codon_start=1 /transl_table=11 /product="GMP REDUCTASE" /protein_id="AAD06382.1" /db_xref="GI:4155370" /translation="MKVFDIEDVQLIPNKIVNSRSCDITVILGKHAFFMPVIPANM QTIINESIAEFLAENGYFIMHFRNGSARIPFVKMKERQLTSSISGVKKECLFVE ELAQGLTPDITIDIAHGSNSVIEMQIKTLPTFVIAVNGVPEAVRELENAG ADATKVGIGPKVCITIKTCFGTGGWQLAALRWCAKARPIIADGGIRTHGDIAS IRFGATVMIGSLFAGHEESGETKIENGIAKYEFGSASEFOKGEKNIEGKKIWIQ HKGSKNTLLEMHQDLQSSISYAGGRDLEAIRKYDYVIVKNSIFNGDTI" complement(7179..7757) /gene="gmba" /note="synonym: jhp0791" complement(7179..7757) /gene="gmba" /note="similar to H. pylori 26695 gene HP0857" /codon_start=1 /transl_table=11 /product="phosphoheptose isomerase" /protein_id="AAD06387.1" /db_xref="GI:4155355" /translation="MIDDLIQEFLAHEALEKESLESQDALQKSVHLLIETLENQOG ILICGNGSASDAQHFAALTGVRKLGSLVLSLTDTSALTAIANDYGEVVFAR QVEALGNKEDILIGISTSGSKNVLKAYEKAKDGLMKTSLAGDGGCKMKFLSDIALI VPSGDTFRIQEMHILMIHLICDCIERHARKN" complement(7750..9141) /gene="waaE" /note="synonym: jhp0792" complement(7750..9141)
gene	
CDS	

```

/gene="waaB"
/notes="similar to H. pylori 26695 gene HP0858"
/codon_start=1
/transl_table=1
/product="putative ADP-D-glycero-D-mannoheptose synthase"
/protein_id="AAD06368.1"
/db_xref="GI:4155356"
/translation="MKKILVIGDLIADYLVWKSERLSPAPVPVLEVKESKNLGGAA
ANVANNITSLKAKVFLGVGDDLEGGHPTSLTKRGIDFTSGVLIDKTRCTTLKTRII
AQNQIVRVDEIKDPLNADLRKNLDPFAEKIQEIDGVILSYDNGVLDLFTQRII
TLANKHKLILCDPKGXDYSKSHASLITENRAELQALHKLKDSHANLSKALQIQE
TYHAMPVLSVGGGIAFLKGBELVNCPTIAKSVYDVTGAGDVIASLTLISLESKL
KEACFANAAAVVQVGMGASALASLEBIALILNQHPLPLEKLETLERNOQKIVF
TNGCFDILHKGHASYLQAKALGDIILVGLNSDINSIKRLKGRPTIVSKDRAPFLAS
LSCVDYVVVFGEDETPIKLIQALKPDILVKADYLNKEVIGSELAKETRLIEPEGYST
SALIEKIKRTHND"
complement(9138..10127)
/gene="gmdH"
/notes="synonym: jhp0793"
complement(9138..10127)
/gene="gmdH"
/notes="similar to H. pylori 26695 gene HP0859"
/codon_start=1
/transl_table=1
/product="ADP-L-GLYCERO-D-MANNOHEPTOSE-6-EPIMERASE"
/protein_id="AAD06369.1"
/db_xref="GI:4155357"
/translation="WRVYIDGLENQITLITGGAGFVGSNLAIFYQENHPKAKVWVLDK
FRSNTLSNNRPSLGHFKLIGFKGEVILADINPLDLRLKFLDYLFHQAAVSD
TTMLNQELVMKTYQAFNLLELARSKKAKVIYASSAGVYGNKAFNIVGSNESPENV

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
3.34e-21	485.00	19549	181
Percent Similarity:	38.04%	Conservative:	94
Best Local Similarity:	25.03%	Mismatches:	222
Query Match:	18.75%	Indels:	226
DB:	1	Gaps:	22

US-10-009-823A-1 (1-502) x AE001509 (1-19549)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 19415 ATGCTTAGGCTTTATGCTGTGTGTCATGATGGATGCAAGCCACCAAAATCGCTTTGGAT 19356

QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 19355 ATTCAGAGTATCAATATATGCGAAGTAAACCACTGGGTTTAAAGTATTCAGGGCTTCT 19296

QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGlySer 56
 DB 19295 TTTGTGATATGCTCTCTCAAGTCAAACTCATGCTACCCACCCATATAAAACGGGTG 19236

QY 57 GlnGlyProAsn-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 DB 19235 GCGGGGAGAGTATGCTTTCTGTGGGCTTGGGTAGCGGTGGATGCGAGCTAAATC 19176

QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys 94
 DB 19175 TTTTCGCAAGCAATATCCAAACACCGATGTCAAACCCGATTCAGGATTCAGGCGAT 19116

QY 95 GlyPheGlnValThrLeuGluAspLysVal-----HisTyrThrArgAlaGlyAsn 112
 DB 19115 GCGTTTTTATCATTCAGCTTATAGGGGATCAGCGCAATTCACAGAGATGGGAG 19056

QY 113 PheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMet----- 130
 DB 19055 TTTCTTTTGTACTGCAAGGAGTTTGTGTACCAACCGGTGGCTTGTGGTGCAGGGTGG 18996

QY 131 -----GlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
 DB 18995 GTGAGAAACGGAGCGATACCGGTAAATAAGGAGCGATACGAGCGCTTTAAAGTGGAT 18936

QY 144 ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLys 163
 DB 18935 AATACCGGCCCTTTAGAAAACATTAGGATTCACGAGTAATGGTG---ATGCCTGTCTAGA 18879

QY 164 ThrSerThr-----AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
 DB 18878 GCGAGTAAACCGCATTTCTATGAGGCGCAATTTAAACGCTGGAAGCATGCGGATCAACA 18819

QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuGluSerTyrLys-----GlyAsnGly 199
 DB 18818 GCGGGG-----GTATTCGCTTAGATTCCTTCAGCCAAACCCCTTCAGATGGC 18771

QY 200 ThrProPheIleSerThrSerAsnTyrSerTyrAlaGlnProMetArg----- 215
 DB 18770 ATTAATCCAGTGTATGATTCAGGCACGAATCTAGCTCAAGTCGCCGAAGACATGGGATCT 18711

QY 216 ValTyrAspGlnGlnGlyAsn-----SerHisAspIleThrValTyr 229
 DB 18710 TTATACAATGAGATGCGGACGCTCTTTTATTGAATGAACCAAGGATTTGGGTGAGC 18651

QY 230 PheAspGlyAla-----ProSer----- 235
 DB 18650 TATAAGAGTGCAGAAATGGTTAAAGACATCCTCCTCTTCGAGAAACACAGCAGCTTCAG 18591

QY 236 -----SerThrGlySerLysThrPheGluTyr 244
 DB 18590 TTGAATGGGTAAAGATTTCTTCAGAACGATTCAGCGGTGAGCAGACTTCAAGCTTA 18531

QY 245 LeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly-----Thr 260
 DB 18530 GTGGCGGCTAAATATCGCATCAATCGGTTAAAGCCAAACAGCAGCTTGAAGCTTATTTG 18471

QY 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 DB 18470 GACGCAAGCAATTCGCTTTGGAAGAACACCAACGAATTAGACGCGGATGAAGACCTTAA 18411

QY 281 Asn----- 281
 DB 18410 AACATTGGTCACTCAAGCCGCGGACCGGTCGCTTAACCTTTTAGCGGCGATAAA 18351

QY 282 ---MetThrAlaPhe-----ThrProThrGlySerAlaThrLysAspLeuAsn 296
 DB 18350 GATGTACGGCTTTCAAAATACATGATATACGCAATCTATTATGCTTAATGCGGATTTGGG 18291

QY 297 AlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheVal--- 315
 DB 18290 CAGTTTAGACCACTGAAGACTTGGCGCCTTAATCCAGCATGACGCTAATATCGTTAAA 18231

QY 316 -----GlyAlaGlyIleGlnProLeuThrLeu 324
 DB 18230 GATCCTAGTTTAGCGGACAATATCAAGATTCAGCGCTTCATATAGGGGTTCACATCAAC 18171

QY 325 AspPheGly-----IleLysSerGlnGln 332
 DB 18170 CAATACGCGATGTTTGAATCAACATAAGATTAATAAATGTCATTAAGAAAAATCTC 18111

QY 333 AsnMetTyr----- 335
 DB 18110 AATATCTTTGTAGTGGGTATTTCTTCAGACGCGTAACGAACATGTTTGTTTAAAAAC 18051

QY 336 -----AlaGlyAlaProAlaSerAlaAla 343
 DB 18050 GCCATGAAGTCTTAATACCGCTTCTTTAATGAAGGGGAGCGTCAGCTAGCAGTTCT 17991

QY 344 -----AlaIleGlyThr--- 347
 DB 17990 AAATTCACCCAGCTACCCATCGCAAGCATTGATGTATAGACAGTTTATAGGCACTAAA 17931

QY 347 ----- 347
 DB 17930 CACGCCATGCGCATTGAGTTTTTATAGGAGTGGGGGGGGAATGGAATTTTAGAGTGATC 17871

QY 347 ----- 347

```

Db 17870 GTGCTGAGCTGGGGAATTAGTAGGGGGTCTAGCGGCTAGGCTAATGTGTGAAGGG 17811
Qy 348 -----AspIleGlyLeuProSerMetMetPro----- 357
Db 17810 GCGCGTTGTCATTTTATAACGCGATCGCTTCGGGCGATGAACCGCTCTTTTGCAA 17751
Qy 358 -----IleGlnThrSerSerGlyAsnSer 365
Db 17750 TTTGACCTAAAATGCGGTGATGCCCCCAACGATCAATTTAGCTTTTGGTTCTCTCA 17691
Qy 366 ThrAlaArgAsnGlySerSerThrArgArgTyxSer-----Gln 379
Db 17690 GGGAGCTTTGACGGATTGACAGCGGTGATAGATTCTTGAACTTATCGGATTGACAA 17631
Qy 380 AspGlyTyxProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGln 399
Db 17630 AACGGCTATCAAGCGGCGGATTTGATGGATGTCGCTTTGATTCTGATGGGTGCTTTTA 17571
Qy 400 GlyLysTyxSerAsnSerGlnValAspPheTyxAsnIleProLeuAlaArgPheThr 419
Db 17570 GGGCGCTTCAGTATGGCAGGACTTTAGCGCTCGCTCAAGTGGCTTTAGCGAATTCGCT 17511
Qy 420 SerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyxSerAlaThrLeuAspSerGly 439
Db 17510 AACGATCGCGGCTTACAGGCTTTGGAGGCAATGTCTTTTCTCAAAACCGAAACTCTGG 17451
Qy 440 GlyProGluPheGlyLeuProGlyThrSerAsnTyxGlyLysLeuSerValAsnGlnLeu 459
Db 17450 CAAGCCTTAATCGTGGCGTAATACGGCGTAGGGGTTCAATTTACAGATCCAAACTA 17391
Qy 460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGly 479
Db 17390 GAGCTAGTAAATGCGATTGACCGCGAGTTTAAACGAATTGATTGCTGTTCAAGGGGG 17331
Qy 480 PheGlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGlu 499
Db 17330 TTTCAACGCGAACTCTAAAGCGGTAAACCATCCGATCAAAATCCTTAACCCCTATTGA 17271
Qy 500 LeuLysArg 502
Db 17270 CTCAGCAA 17262

RESULT 32
AC095028
LOCUS
DEFINITION
Leishmania major chromosome Unknown clone P864 strain Friedlin, ***
SEQUENCING IN PROGRESS ***, 4 ordered pieces.
AC095028
AC095028.2 GI:117352423
HTG; HTGS_PHASE2.
Leishmania major
Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 90250)
Myler,P.J., Sisk,E., Cawthra,J., Nelson,S., Vogt,C., Robertson,L.,
McDonagh,P., Ivens,A., Nguyen,D., Munden,H., Stuart,K., Mack,J.,
Marty,A., Rinta,J. and Seyler,A.
Direct Submission
Submitted (15-SEP-2001) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
On Dec 5, 2001 this sequence version replaced gi:15624885.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 7210: contig of 7210 bp in length

```

FEATURES

source

```

* 7211 7411: gap of unknown length
* 7412 27989: contig of 20578 bp in length
* 27990 28189: gap of unknown length
* 28190 58975: contig of 30786 bp in length
* 58976 59175: gap of unknown length
* 59176 90250: contig of 31075 bp in length.

```

Location/Qualifiers

1..90250

/organism="Leishmania major"

/mol_type="genomic DNA"

/strain="Friedlin"

/db_xref="taxon:5664"

/chromosome="Unknown"

/clone="P864"

/note="This PAC appears to contain bacterial DNA. It is not yet clear whether it is derived from a contaminant or a symbiont."

ORIGIN

Alignment Scores:

```

Pred. No.: 2 05e-20 Length: 90250
Score: 485.00 Matches: 153
Percent Similarity: 39.82% Conservatives: 74
Best Local Similarity: 26.84% Mismatches: 177
Query Match: 18.75% Indels: 166
DB: 2 Gaps: 19

```

US-10-009-823A-1 (1-502) x AC095028 (1-90250)

```

Qy 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla 27
Db 80344 GGAGTCAGCGGCATCAATGCCCGGCACAGAAATCGGACGTGATCGGCAACACATCGCC 80403
Qy 28 AsnAlaSerThrIleGlyTyxLysGlnGlnValValPheGlnAspLeuPheSerGln 47
Db 80404 AACGGAAACACGTCGGCTTCAAGCAATCGACCGCGAGTTTCGCCGAGCTCTACGCC--- 80460
Qy 48 AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln 67
Db 80461 -----GGCTCG-----AAGGTCGGCTGGGCACGGC 80487
Qy 68 ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr 87
Db 80488 GTGAACTCGGTGGTGCGATGTTTCCAGCAGGCGCAATATCGAGCGGAGCGCGGCTCGCTG 80547
Qy 88 AspLeuAlaIle---GlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysVal 105
Db 80548 GACGTCGCGATCACCGGGGCAACGGCTTCTTCGCGATGTTTCGCCACCGCGCTCGGTG 80607
Qy 106 HisTyxThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSer 125
Db 80608 TTCTTCTCGGCAACGGCGCATTTCTCTGCGCACCGACGCGCGCGCATCACCAACGCGCAG 80667
Qy 126 GlyPheThrLeuMetGly----- 131
Db 80668 GCGCTGAGCTGACCGGCTATCCGGCGGGGTGCGCGGGCGGCGGCTGCGACCGCAA 80727
Qy 132 -----SerArgIle----- 134
Db 80728 CCGCTGGTGGTGTCCAGCGCCAGATGCGCGCGAAGGACCGACCGCATCACCGCGCAA 80787
Qy 135 -----SerAsnAsnProAsnIleLysLysGlu----- 143
Db 80788 TACAACCTGGACTCGCGCGCGGCTTCTCGCAATCCAGCGTGCAGCGCAACCGCGCTAAT 80847
Qy 144 -----ThrLeuGluProIle--- 148
Db 80848 ACGGGATCAGTTCGGTACAGTTCGCGAGTTTCGACGCAACCGCGCTGGAGGACTTCACC 80907
Qy 149 ---GlnLeuAspPheAsn-----AspProThrValAlaLysSerProAlaLysThrSer 165
Db 80908 TACCAGCTGTCTTTCGATGGCACCGGATTAACACCTGACCCGCTTCTGTATGGGCAACGGT 80967

```

```
Qy 166 ThrAlaLeuAsnAlaVal-----Val 172
|||
Db 80968 ACTCAGCGGGCGACACCTTTCAGATAGACGGCTGACGCTGACACGGCGCGGGTATG 81027
|||
Qy 173 AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeu 192
|||
Db 81028 AACACCGTGCACCC-----TTTACCGTGCATCGCTCACCCTTAACCTC 81072
|||
Qy 193 GluSerTrpLysGlyAsnGlyThrProPheLysThrSerAsnTyrSerTyrAlaGln 212
|||
Db 81073 ACTAGTCGCGCGGCAAT-----CCCGCTGCAGCAGCAAGCTTCGACTACACGAC 81123
|||
Qy 213 ProMetArgValTyrAspGlnGlnGlnGlnHisAspLeuThrValTyrPheAspGly 232
|||
Db 81124 ACGATCAGGTGTTCGACTCGCTCGGCAATGCCAAGCAGGTGACGCTATTCGTCAG 81183
|||
Qy 233 AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGlu 252
|||
Db 81184 GGTCCGACCGCGGGAACGCTAGCGCTGGGATGTCTACGTG----- 81225
|||
Qy 253 AspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly-----Thr 270
|||
Db 81226 -----ACGACAGCAGCGCGCGCCCACTGTCGCTCCCAATGAG 81264
|||
Qy 271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
|||
Db 81265 ATGTGTTTAACCCTGAACGGCGCGCTG-----GTAACCTCCGCGCGCGCG 81309
|||
Qy 291 AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPhe 310
|||
Db 81310 GCAACG----- 81315
|||
Qy 311 SerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSer 330
|||
Db 81316 -----ATCCCTGTGACCGTC----- 81330
|||
Qy 331 GlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGly 350
|||
Db 81331 -----CCCGAGTCCGCGCGCGCGCGCG----- 81354
|||
Qy 351 LysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGly 370
|||
Db 81355 -----ATGAACATCCACTGCAGCTGCGCGCGCGCGCGCGCGCGCGCG 81399
|||
Qy 371 SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal 390
|||
Db 81400 GACACGATCCGGAAGTCGATCGTCAGACGCGCTATACCTCGGCTCGCTGCGGCTTC 81459
|||
Qy 391 ThrIleThrSerGluGlyLysLeuGlnGlyLysThrSerGlnValValAspPhe 410
|||
Db 81460 TCGATCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81519
|||
Qy 411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGlyAsnAsn 430
|||
Db 81520 GGTGAGTCCGATGAGCTCGTCTCGCAACGCTCGAAGGGCTCAAGCCGAGCGCGCAAC 81579
|||
Qy 431 HisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
|||
Db 81580 GTCTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81636
|||
Qy 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
|||
Db 81637 ATGGCCAGCTGAGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81696
|||
Qy 471 ValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
|||
Db 81697 GTCAACCTGATCATTCGCCGCGCAACTACGAGCGCAATCGCGAGACCATCCAGCGCGCAG 81756
|||
Qy 491 AspThrMetLeuGlnLysAlaLeuGluLeu 500
|||
Db 81757 GACAGGTGCTGCAGACGCTGTCACATC 81786
|||
```

RESULT 33

```
AL591976
LOCUS
DEFINITION
ACCESSION
AL591976
VERSION
AL591976.1
KEYWORDS
SOURCE
ORGANISM
Listeria monocytogenes
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
AUTHORS
Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,
Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,
Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P.,
Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
Dussutget, O., Etian, K.D., Fsihi, H., Portillo, F.G., Garrido, P.,
Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J.,
Jackson, D., Jones, L.M., Kaerst, U., Kref, J., Kuhn, M., Kunst, F.,
Kurapat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E.,
Nedjari, H., Nordisk, G., Novella, S., de Pablo, B., Perez-Diaz, J.C.,
Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N.,
Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J. and Cossart, P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
21537279
TITLE
MEDLINE
11679669
PUBMED
2 (bases 1 to 250050)
REFERENCE
Glaser, P., Frangeul, L. and Rusniok, C.
Direct Submission
AUTHORS
TITLE
JOURNAL
Microorganisms Pathogenes, 25 rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE
COMMENT
E-mail: pglaser@pasteur.fr
Phone: +33 1 45 68 89 96, Fax: +33 (0) 1 45 68 87 86.
FEATURES
Location/Qualifiers
1..250050
/organism="Listeria monocytogenes"
/mol_type="genomic DNA"
/strain="SGD-e"
/db_xref="taxon:1639"
complement (116..121)
/gene="lmo0598"
complement (116..121)
/gene="lmo0598"
join (287..292,300..632)
/gene="lmo0599"
/gene="lmo0599"
/gene="lmo0599"
/gene="lmo0599"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98677.1"
/db_xref="GI:16409988"
/db_xref="TrEMBL:Q92E53"
/translation="MEVNPQKKGVLLELCLFLQKDCGYELANOVSKYIEVAGEA
IYVLRILRVKEBYCSTYLVESNGPSRKQTLTKGBIYLNELISEWNFTDSYAKLL
TEGEAVNE"
610..615
/gene="lmo0600"
610..615
/gene="lmo0600"
625..1215
/gene="lmo0600"
625..1215
/gene="lmo0600"
625..1215
/transl_table=11
/protein_id="CAC98678.1"
/db_xref="GI:16409989"
/db_xref="TrEMBL:Q8Y9D2"
/translation="MNQDFLNLNQLRLELDPKERRELLSDYQEHFRNGIEAGKSEE
QIVFDLGPPEIAADIISERGIREPTMDYVYVPRKNQENRSVSKILIGVGLFL
```



```
DVCLIPIMVSLVSLVSWTVVGAFLPSVLGVGLFGADPFYQMFVSGVLGLG
LMLLFVANALQTLTSKVTWAILQWHKYAVKGGKNA"
/gene="lmc0601"
RBS 1193..1198
/gene="lmc0601"
1193..1198
/gene="lmc0601"
1208..2308
/gene="lmc0601"
1208..2308
/gene="lmc0601"
1208..2308
/gene="lmc0601"
1208..2308
/notes="similar to cell surface protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98679.1"
/db_xref="GI:16409990"
/db_xref="EMBL:Q8Y9D1"
/translation="MKHKKSLKFLFAGLVLFIIIGAVFMTMTGKMEKEPLTKQ
WLSITENIKKIFASERDASIEWKSTTKYIELKGNYSANAKAIQOLPVSVDGT
SPDITVPEDDNYNGFKIYAKQKQVLYITKDKLADLEVKSHSGIDIVADFKVK
FVSTNSGKLVNLEANTAOATSGDGLTSLNKAANGSIETDSKGTSLTNLTGLEY
NGSGDNYVAGVAKKLTAKIDSGLTSGVTVDLAVITTSGGDIDANTKGICQARS
DGSSELAGATNNVTKTSSGDIIDVAFKQVKNIEINSDSGEVELELPQDFKAIYAS
SNQSVKAPTSDSNDNRVTVKTSDDIKIEK"
join(2400..2405,2414..2914)
/gene="lmc0602"
RBS 2400..2405
/gene="lmc0602"
2414..2914
/gene="lmc0602"
2414..2914
/notes="weakly similar to transcription regulator"
/codon_start=1
/transl_table=11
/protein_id="CAC98680.1"
/db_xref="GI:16409991"
/db_xref="GOA:Q8Y9D0"
/db_xref="EMBL:Q8Y9D0"
/translation="MDFHKKATNSDAEAIQHVATSWHTTYQDLIPGDVQDDFLERF
YNETLHNRISATPEAVQADKIVGFANFIELEKKSLEAAFYLLPVTQKGLGTEL
LEYGMTLFVPLVPMFVYVNEKGNETAHFYKAKGVQVEFTEDFYPLETIRFNLNH
HAPEEP"
complement(2909..2956)
/gene="lmc0603"
terminator complement(2909..2956)
/gene="lmc0603"
2920..2967
/gene="lmc0602"
2920..2967
/gene="lmc0602"
2920..2967
complement(join(2962..3348,3356..3361))
/gene="lmc0603"
complement(2962..3348)
/gene="lmc0603"
/codon_start=1
/transl_table=11
/protein_id="CAC98681.1"
/db_xref="GI:16409992"
/db_xref="EMBL:Q8Y9C9"
/translation="MRPIIAYUSIFVLGIFSAALLVETIYDNTVQLVFSAILFAAPV
ILVATIGIFYGFKASYFTFAIMGFAYGVAIVILSIQVSGMLISVGSVILAG
VIMALLAIIPFFLRGGKSTSGKAATK"
complement(3356..3361)
/gene="lmc0603"
complement(3400..3761)
/gene="lmc0604"
complement(3400..3744)
/gene="lmc0604"
/notes="similar to B. subtilis YvIA protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98682.1"
/db_xref="GI:16409993"
/db_xref="EMBL:Q8Y9C8"
/translation="MLDKKIISALYSFSLFFAFVITPAIINWCTTKDKIRHHVKNALL
```

```
TOATFVAGIVVMILLVNNVPFSTNSADTINFLATVFFVFLNVSILIFNLIRGTR
IVKSHDDNARC"
complement(3756..3761)
/gene="lmc0604"
RBS complement(3891..5244)
/gene="lmc0605"
complement(3891..5231)
/gene="lmc0605"
/notes="conserved hypothetical membrane protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98683.1"
/db_xref="GI:16409994"
/db_xref="GOA:Q8Y9C7"
/db_xref="EMBL:Q8Y9C7"
/translation="MSSVHKVASISLFTLAWPIFLFQFLRLMISYIDVFMGLHYSDDA
VATCAVNAQLIVISIIYGFISVGVQIIIVAOIGAKKKHEIENVITNGIYVATLIGIV
MSIIFPMSKNFLTMGIDPHLVQVGAPFIEIIGSSVVAIHASILPILRAKGYVRQ
SILVPTISINNVGNLYFLYGLYLDYGVAGVISTAVNFVGMGLAIWMLRKYIG
YTHFKLEQVSKKLYSILRLGPSAGENSIVAGSOLVVTAIIALIGTEALTTKVYA
STVQPVAFALAGQASQIIIGRAVGAKEDIKAYQGLRSWKIGLVVAIVVSYSIYL
FABPIMLFTTNTTEIIAMTKELFLLSIFLELRATNIIIISSLNSTGDVRFPTCGLI
VMIVSPFYSVLGISAGLVGLWLAIIIDEGVRAVLMYRRRSKWSLKSVI"
complement(5239..5244)
/gene="lmc0605"
RBS 5384..5858
/gene="lmc0606"
5384..5389
/gene="lmc0606"
5397..5858
/gene="lmc0606"
/notes="similar to transcription regulator Marr family"
/codon_start=1
/transl_table=11
/protein_id="CAC98684.1"
/db_xref="GI:16409995"
/db_xref="GOA:Q8Y9C6"
/db_xref="EMBL:Q8Y9C6"
/translation="MSSKNQDLGHSVTKAFNFKHAIKSFQIPGYKSETRFFILS
RGKSRGPKIRVSDLGHMLRIKPSVTQMTQSLGKGLIKRVQNPDKRSMYVELTEV
GAGVSEKMLDFEFAQSFEDQFLEJGDDMKLLITLLEKLTLYLNTKSENKEA"
join(5865..5871,5878..7590)
/gene="lmc0607"
RBS 5865..5871
/gene="lmc0607"
5878..7590
/gene="lmc0607"
/notes="similar to ABC transporter, ATP-binding protein"
/codon_start=1
/transl_table=11
/protein_id="CAC98685.1"
/db_xref="GI:16409996"
/db_xref="GOA:Q8Y9C5"
/db_xref="EMBL:Q8Y9C5"
/translation="MKRLKPYWLSITAVLVLTGQVIGQLYPLTMSNIIDKGVVTGD
```

Alignment Scores:

Pred. No.:	7,38e-20	Length:	250050
Score:	484.50	Matches:	137
Percent Similarity:	42.21%	Conservative:	77
Best Local Similarity:	27.02%	Mismatches:	188
Query Match:	18.74%	Indels:	105
DB:	1	Gaps:	11

US-10-009-823A-1 (1-502) x AL5191976 (1-250050)

QY	1	MetMetGlySerLeuPheIleGlyAlaThryGlyMetIysThrHisSerThrGlyLeuGly	20
Db	90516	ATGAATCAACTATGTATACAGCTATTTCGGATGAATCGTTCACACAGCATATCA	90575
QY	21	ThrValSerAsnAsnIleAlaAsnAlasThrIleGlyTyrLysGlnGlnValVal	40
Db	90576	GTAACATCAATAATATTGCCATGCCACACAGCATATAAACAAGCGTGGT	90635

QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
 Db 90636 TTCAATGATTACTTTTACCAAAACACAAATGGGATCTGTTCCAGCGGACATTTATGCTGGA 90695
 QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
 Db 90696 ACAAAACCAATGAGTTTCGGTTCGGTTCGAAATTCGGCGGATTTTAACGTACTATACA 90755
 QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlySerGlyPhe 96
 Db 90756 GCAGGTTCCCGACACTTCAACTGGCAGAAATAAAGATCGACACTACAAGGCCCGGCTTT 90815
 QY 97 PheGlnVal-----ThrLeuGluAspLysValHisThrThrArgAlaGlyAsnPheArg 114
 Db 90816 TTCAATGCTGTGATAACGCTGGTGGGAATATCGTTTATACGTGACGTACTTTTGA 90875
 QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
 Db 90876 GTATCTGACAACTATTTTAAACACGACGAGGAAATACGTTATGGGTTACGCAACG 90935
 QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
 Db 90936 GACAAAATGTAACGTTTAAACGGTAAATTTACACCCATCCAAATTCGCTAAATAGC 90995
 QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
 Db 90996 GCAATTCAGGGAGCAACAAAAATGGTAGCTAAGCGGTAACTCCACATGATGTTGG 91055
 QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
 Db 91056 GCGCAAAAA---GATACGATTCTTCGAG----- 91082
 QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
 Db 91083 -----CTG 91085
 QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
 Db 91086 TCTGTATACGATAATGCTGGCGGAAAAACATAAACTTCAAGTCAATATGAAGCAGCTACA 91145
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 Db 91146 CCAGATCGGACGGTAATGTTCTACGATATGAATTCAAATG-----GAC 91193
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 Db 91194 GCGAAAGCATTAATCTCTCCA-----GTGACAGGACACTTAATTATAC 91235
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 Db 91236 AATGCGGAGGTGAATACGAAACCCAGACGCA----- 91268
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 Db 91269 ---CTTAAATAATTCAAATCAATTCACAGTAAACGGC----- 91304
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 Db 91305 -----AAACAGTCAAT 91316
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 Db 91317 ATG-----GGCTTAAACCTAAGTGCCTAACCA 91343
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
 Db 91344 AACTAC-----GGAACGACCAAA 91361
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 Db 91362 GTATCTCACCACACTTCTGTATGCAAGGTGTCGACACTGTAAAGATTTATGACGATTACC 91421

QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 Db 91422 GATTCGTGCTATATTCAGTGCAGTATTCAGATGGTACAGTATCCCGAGTGGCCCACTC 91481
 QY 414 ProLeuAlaAaGpPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
 Db 91482 CGGTTGCTACGTTCTCCCAATGAAGACGGTTTACTCAAAATGGGAAACGGCGAATATGTT 91541
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 Db 91542 CCAGATTTATCTTCTGGTGATGCAATACGGCTGCTGCGCAAAATGGCGCTGGCGGA 91601
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 Db 91602 ATTAGTGTCTTCTCCAGAGGATCAACGATAGCTGTCCCGTGAATTCGTTAACTTA 91661
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
 Db 91662 ATGACATATCAAGTGGTTTCCAAAGGAATACAAAGGTTATTCGTGCGCGATGACGTG 91721
 QY 494 LeuGlnLysAlaLeuGluLeu 500
 Db 91722 ATGAAACAAATTGAACTTG 91742
 RESULT 34
 LOCUS AL596166 260050 bp DNA linear BCT 06-JUN-2002
 DEFINITION Listeria innocua Clip11262 complete genome, segment 4/12.
 ACCESSION AL596166 AL592022
 VERSION AL596166.1 GI:16413128
 KEYWORDS
 SOURCE Listeria innocua
 ORGANISM Listeria innocua
 Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 1
 REFERENCE
 AUTHORS Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Delhou, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dusurget, O., Entian, K. D., Fsihi, H., Portillo, F. G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L. M., Kaerst, U., Kref, J., Kuhn, M., Kunst, F., Kurapat, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E., Nedjari, H., Nordisk, G., Novella, S., de Pablo, B., Perez-Diaz, J. C., Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. and Cossart, P.
 TITLE Comparative genomics of Listeria species
 JOURNAL Science 294 (5543), 849-852 (2001)
 MEDLINE 21537279
 PUBMED 11679669
 REFERENCE 2 (bases 1 to 260050)
 AUTHORS Glaser, P., Frangeul, L. and Rusniok, C.
 TITLE Direct Submission
 JOURNAL Microorganisms Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE
 COMMENT E-mail: pglaser@pasteur.fr
 Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
 FEATURES
 source
 1. 260050
 /organism="Listeria innocua"
 /mol_type="genomic DNA"
 /strain="Clip11262"
 /db_xref="taxon:1642"
 331..1225
 /gene="lin0677"
 331..336
 /gene="lin0677"
 344..1225
 /gene="lin0677"
 /note="conserved hypothetical protein"
 /codon_start=1
 /transl_table=11
 gene
 RBS
 CDS

/db_xref="GI:16413138"
/db_xref="GOA:Q92DX5"
/db_xref="TrEMBL:Q92DX5"
/translation="MEFEFLAVIVFVSFVAFVFFFLPFLKGRNIPNSVKVFGMAIS
IPVATGVDSGTTLPOLLRVTSVVFGLAKLVEIIAVIPKMGFMIDYDLGFSQ
VNLDPSGTGNSITAILDFFVIFLSQGMDFLYLMLKSFETFAVSILFEKGF
IDLLGLTGLFALSAVSIALPIMGSI FIVNLVILAFISKSAPQINIPNNAFLIKITGI

Alignment Scores:
Pred. No.: 7,73e-20 Length: 260050
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 1 Gaps: 11

US-10-009-823A-1 (1-502) x AL596166 (1-260050)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
DB 24803 ATGAATCAACTATGATACAGCTATTCTGGGATGAATCGGTCCACAGCAATATCA 24862

QY 21 ThrValSerAsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnValVal 40
DB 24863 GTGACATCAATAATATATGCAATCGCAACACGACGAGGTATATAAACAACAAAGCGTGT 24922

QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
DB 24923 TTCAATGATTTACTTTACCAAAACACAAATGGGATCTGTTGAGCGGACCTTTATGCTGGA 24982

QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
DB 24983 ACAAAACCAATGAGCTTCGGTTCGGTTCGAAATTCGAAATTCGGCAATTTTAAACGATTATACA 25042

QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
DB 25043 GCAGGTTCCTCGCATCACTCACTGGCAGAAACAAAGATGCACACTGCAAGCGCCGCGCTTT 25102

QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
DB 25103 TTCATTGTCGCGATAACGCTGGTGGGAATATCGTTTACACAGACGCGTASCCTTGCA 25162

QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
DB 25163 GTATCTGCATAACTATTATTAACCAACCAACAGGAATAATACGTTATGGGATATGCAACA 25222

QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
DB 25223 GATAAAATGGCAACGTTTAAACGGNAACCTTGCACCAATTCCAATCCCAATTAATAGC 25282

QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
DB 25283 GCATTCGCGGAGCAACAAAAATGGTAGCTTAAGCGGTAACTATCCACTTGATGG 25342

QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSer 194
DB 25343 GCGGAAAAA---GATACGATTCTTCGAG----- 25369

QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
DB 25370 -----CTT 25372

QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
DB 25373 TCTGTATACGATATGCTGTGGGAAACATAAATCAATCAATGAAAGCTGGACA 25432

QY 234 ProSerSerThrGlySerThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
DB 25433 CCAGATCGGATGGCAATGTTCTCTATGATATGAATTCAAATG-----GAC 25480

QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
DB 25481 GGAAAGCGTTAATCTCTCCA-----GTACAGGAACCTTAATATAC 25522

QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 25523 AACGGCAGGCGAACTTACAAACCTGATGCA----- 25555

QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 25556 ---CTTAAACCAATTCCAATCAATTCCTGTTAAACGGC----- 25591

QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
DB 25592 -----AAACAAGTCAAT 25603

QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 25604 ATG-----GGCTTAAACCTTAAGTGGCTTAA 25630

QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
DB 25631 AACTAC-----GGTACCACCAA 25648

QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
DB 25649 GTATTCTCACCACACTTCTGACGGAAGCGCTGCACTGTAAAGAGACTACGCAATTACC 25708

QY 394 SerGluGlyLysLeuGlnGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
DB 25709 GATTCGTGTTATATTCAGTCACTGAGTACTCAGCGTACAGTTATCCAGTTCGCCAATT 25768

QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
DB 25769 GCGGTTGCTCTTCTTCCAAATGAAGCGCTTAGTCAAAATGGGGAATGCGCAATATGTT 25828

QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
DB 25829 CAGGATTATCTTCTGGAGATGCGAGTTTACGCGCTTCTGCGCAAAACGAGCTGGCGGA 25888

QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
DB 25889 ATTAGCGCTCTCTTATTAGAAAGCTCAAAAGTACAGTATTATCCCGCAATTGTTAACTTA 25948

QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
DB 25949 ATGACATACCAAGATGGTTCCTCAAGGCAATACAAAAGTTATCCGTGTTCGAGATGACGTG 26008

QY 494 LeuGlnLysAlaLeuGluLeu 500
DB 26009 ATGAACCAAAATTGTGAATTG 26029

RESULT 35
AE017324 290242 bp DNA linear BCT 30-APR-2004
LOCUS Listeria monocytogenes str. 4b F2365, section 3 of 10 of the
DEFINITION complete genome.
ACCESSION AE017324 AE017262
VERSION AE017324.1 GI:46880047
KEYWORDS Listeria monocytogenes str. 4b F2365
SOURCE Listeria monocytogenes str. 4b F2365
ORGANISM Listeria monocytogenes str. 4b F2365
REFERENCE 1 (bases 1 to 290242)
AUTHORS Nelson, K.E., Fouts, D.E., Mongodin, E.F., Ravel, J., DeBoy, R.T.,
Kolony, J.F., Rasko, D.A., Angiuoli, S., Gill, S.R., Paulsen, I.T.,
Peterson, J.D., White, O., Nelson, W.C., Nierman, W.C., Beanan, M.J.,
Brinkac, L.M., Daugherty, S.C., Dodson, R.J., Durkin, A.S., Madupu, R.,
Hatt, D.H., Selengut, J., Van Aken, S., Khouri, H., Fedorova, N.,
Forberger, H.A., Tran, B., Kathariou, S., Wonderling, L.D.,
Uhlrich, G.A., Bayles, D.O., Luchansky, J.B. and Fraser, C.M.
Whole genome comparisons of serotype 4b and 1/2a strains of the
food-borne pathogen *Listeria monocytogenes* reveal new insights into
the core genome components of this species
Nucleic Acids Res. 32 (8), 2386-2395 (2004)
JOURNAL PUBMED 15115801

3796...-4989

/locus tag="LMOf2365_0569"

/note="identified by similarity to OMNI.NTL01BS01695;

match to protein family HMM PF00144"

/codon_start=1

/transl_table=11

/product="penicillin-binding protein, putative"

/protein_id="AAO703351.1"

/db_xref="GI:46880051"

/translation="MNRERERKHKRKNMIIVLSTFLLFIATSWMIYEKSKQEOIV
SAPTKKKATATNPENNAANDEKETPEPVEPKETIVKQEDINVLQIKFRSGS
LAALVADGDKTIQKGWYANRDKEVLNTPDILFYIGSSQKAIATALLOLEEKGININ
WDYSIYPDFNGNKLVIPFNMTHTSIGLVRKPULSNNTPDQIKALKERKGNISQKQK
DDTIDANTAVGVYLVERKVSQGLATYIQENITFPAGMKGTGYQKQDVGGKNGISQKQK
IDENGVFQSPKLADIUKLYGAGDISMTSDMYLPFKALMDPKKIISEASLKAMFTSGK
ITVGWGFGFYDPGSYNHSGVDNSVSHTGRTYYVLFNNQNHIASFCKVNNDIY
GFENK"

/complement(5030..5950)

/locus_tag="LMOf2365_0570"

/complement(5030..5950)

/locus_tag="LMOf2365_0570"

/note="identified by match to protein family HMM PF01497"

/codon_start=1

/transl_table=11

/product="ABC transporter, substrate-binding protein"

/protein_id="AAR03352.1"

/db_xref="GI:46880052"

/translation="MYQRKWAAILMTVLVLAVLTACGSSDDKDSKEKESAKTVTD
IMRWKVEPTTPKRIVLQNVSEMLGVKPGVTDIYITTEPATKTGETSVGNKDES
IEIKLANLEDLLIIDSYQDLLENELGVKVPYITFHGDTPDKQLSNLIANKKAERE
KWDOYASSKEKATLADKAVANEKAAVIOFYGKEIVVHDKAVFDGLYSGAGTPTD
AKANKETKAISSAIPYAGADELFLIMPADGNTSDVDEMLKGVWKDIPAVKKNQV
KYVDNTKSDSYDAIPLQCMEDAQQITGK"

/complement(6061..6411)

/gene="srlB"

/locus_tag="LMOf2365_0571"

/complement(6061..6411)

/gene="srlB"

/locus_tag="LMOf2365_0571"

/EC_number="2.7.1.69"

/note="identified by similarity to SP:P05706; match to
protein family HMM PF03829"

/codon_start=1

/transl_table=11

/product="PTS system, glucitol/sorbitol-specific, IIA
component"

/protein_id="AAO703353.1"

/db_xref="GI:46880053"

/translation="WTIKIMEIGLVPFAFBEKVIILFGPTATNEIRLSIVHEHQFS
VPNALTKGNTLLIGQAYTIEBGSDANKNLIELGHISVYFRSQNEVLPGAIVSE
EIPPTITVDSIQF"

/complement(6430..7416)

/gene="srlE"

/locus_tag="LMOf2365_0572"

/complement(6430..7416)

/gene="srlE"

/locus_tag="LMOf2365_0572"

/EC_number="2.7.1.69"

/note="identified by similarity to GP:4928286; match to
protein family HMM PF03612"

/codon_start=1

/transl_table=11

/product="PTS system, glucitol/sorbitol-specific, IIBC
component"

/protein_id="AAO703354.1"

/db_xref="GI:46880054"

/translation="MTQSJHNKGGGWGVLDTTPSRPKVVISITGGGTHPVALKIK
AEUSAGADVDFNKSPIEEEMCVVIDCGGTARIQVIFEWKPIPIVDVLPPSSGPLAN
HITEDIYVSGVKEDIEVDRDTRTEKPFPRDEKFDYAKIKDOERASKUEGLFVLR
FSRIGKAWGVFQAQRSVDMLNIPFMAFISMIIINTYGKLIKANTLSPLA
GSLWGMILLIICSLFPLSPGPAVIAQVILGVLSIQGIALGPVTPVALPALFAIN
AQAGCDPIPVGISLAEARKTQVQFGVPALLYSERMITGVLAIVVAWFVFIGMY"
AACCDPIPVGISLAEARKTQVQFGVPALLYSERMITGVLAIVVAWFVFIGMY"

/complement(7437..7958)

CDS	/gene="srla"		/locus_tag="LMOF2365_0573"		complement(7437..7958)		163757 GACAAAATGCGCAACGTTTAAACGGTAATTTGCAACCAATTCAAATCCCGCTAAATAGC		163816	
	/gene="srla"		/locus_tag="LMOF2365_0573"		/note="identified by similarity to SP:P56579; match to protein family HMM PF03608"		155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu		174	
gene	/codon_start=1		/transl_table=11		/product="PTS system, glucitol/sorbitol-specific, IIC component"		163817 GCAATTCAGGAGAACCAAAAAACGGTACTTAAAGCGGTAAACATCCACTAGACTGG		163876	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer		194	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		163877 GCGGAAAAA---GATACGATTCTTCCGAG--		163903	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		195 TrpLysGlyAsnGlyThrProProLysSerThrSerAsnTyrSerTyrAlaGlnProMet		214	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		163904 -----		163906	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla		233	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		163907 TCTGTATATGATAATGCTGCGGGAACATATAACTTCAAGTGAATATGAAAGCTGCATACA		163966	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp		253	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		163967 CCAGATGCGAGCGGTAAATGTTTCTCTACGAATATGAAATTCAAATG-----GAC		164014	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe		273	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		164015 GGAAGACATTAACTCCCTCCA-----GTACAGGAGCACTTAATTATC		164056	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys		293	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		164057 AATGCGCAGCGGAATTAAACAAACCCAGACCA-----		164089	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		294 AspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn		313	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		164090 ---CTTAAAAATATCCAAATCAATTCACAGTAAACGCGC-----		164125	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn		333	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		164126 -----AACAACTCAAT		164137	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro		353	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		164138 ATG-----GGCTTAAACCTAAGTGGCTTAACC		164164	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer		373	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		164165 AACTAC-----GGAACAAACCAA		164182	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr		393	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		164183 GTATTCTCAACCACTCTGACGGTAAAGCGCTGCGACTGTAAAAGATTATGCGATTACC		164242	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle		413	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		164243 GATTCTGGCTATATTGAGTGGTACTGACATGGTACAGCTTATCCAGTGTGCCCACTT		164302	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer		433	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		164303 CGCGTGGCTACTTCTTCCATGAAGACGGCTTAGTCAAAATGGGAACGGCGAATATGTT		164362	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys		453	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		164363 CCAGGATTATCTCTGCGCATGACGATATACGGGCTGCTGCGCCAAATGGCGCTGGCGGA		164422	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet		473	
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		164423 ATTAGCGGTCTTCTTAGAAGGTTCAACGTAGACTTGTCCCGCGAATTCGTTACTTAA		164482	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"		474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet		493	
gene	/codon_start=1		/transl_table=11		/product="conserved hypothetical protein"		164483 ATGCATACCAAAAGTGTCTTCCAGGGAATACAAAAGTTATTCTGTGGCGGATGACGTG		164542	
	/protein_id="AA03355.1"		/db_xref="GI:46880056"		/translation="MNVLVLSLFIATQGLAFQVRYOHHMKNVADKYAEKTYVL YSEMERLKFTSVALVNVNENVEHCILTKTGFAPKTFNTYHKKDLSELIDL SSRKNVITQKAIKIVNSCKWL"		494 LeuGlnLysAlaLeuGluLeu 500			
CDS	/locus_tag="LMOF2365_0574"		complement(7983..8363)		/locus_tag="LMOF2365_0574"		164543 ATGAACAAATTTGGAACCTTG		164563	
	/gene="srla"		/locus_tag="LMOF2365_0574"		/note="identified by similarity to OMV1.NTLO1LI0545"					

Alignment Scores:

Pred. No.: 8.e-20 Length: 290242
 Score: 484.50 Matches: 137
 Percent Similarity: 42.21% Conservative: 77
 Best Local Similarity: 27.02% Mismatches: 188
 Query Match: 18.74% Indels: 105
 DB: 1 Gaps: 11

US-10-009-823A-1 (1-502) x AE017324 (1-290242)

QY	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
DB	163337	ATGAATCAACTATGATACAGCTATTCTGGGATGAATCGTCCACAGCAATTATCG	163396
QY	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal	40
DB	163397	GTAACATCAATAATATTGCCAATGCCACACAGGATACAAAAACAAACGTCGTT	163456
QY	41	PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer	56
DB	163457	TTCATGATTACTTTTACCAAAAACAAATGGGATCTGTTCAGCGCGACTTATGCTGA	163516
QY	57	GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr	76
DB	163517	ACAAACCAATGAGTTTCGGTTCGGTTTCAAAATTTGGAGCGATTATACCGATTATACA	163576
QY	77	GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe	96
DB	163577	GCGGGTTCGCCAACACGACGCGGACAGAAATAAGATGCAGCACTACAAGTGGCGCTTT	163636
QY	97	PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg	114
DB	163637	TTCAATCGCGGGGATATGACAGGGGCAATATCGTTTACCGGTGACGGAGCTTCGCC	163696
QY	115	PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgile	134
DB	163697	GTTTCCGACACCAATATTATTTAATACCAACAGGGAATAAGTATGCGTTACGCAACA	163756
QY	135	SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp	154

RESULT 36
AX417038
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
SCIENTIFIC (CNS) (FR)
LOCATION/Qualifiers
1..349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"
/note="seq 10 splitted into-seq 4029; from 300001 to 649980-seq 4030; from 600001 to 949980-seq 4031; from 900000 to 1163020"

ORIGIN
Alignment Scores:
Pred. No.: 1,1e-19 Length: 349980
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-1 (1-502) x AX417038 (1-349980)
QY 1 MetMetGlySerLeuPheLeuAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
DB 339538 ATGAATCAACTATGTATACAGCTATTCTGGGATGATCGGTTCCAAACAGCATTTATCA 339597
QY 21 ThrValSerAsnAsnAlaAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
DB 339598 GTGACATCAATAATATTGCCAATGCGAACACGACGCGGTTATTAACAAACGCGCTT 339657
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
DB 339658 TTCAATGATTTTACCTTACCAAAACACAAATGGGATCTGTCAGCGGCACTTTATGCTGGA 339717
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
DB 339718 ACAACCCCAATGAGCTCGGTTCCGGTTCGAAATTTGGGCAATTTTAAACCGATTATACA 339777
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
DB 339778 GCAGGTTCTCCGACATCAACTGGCGAGAACAAAGATGCACACTGCAGCGCGCGCTTT 339837
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
DB 339838 TTCAATGCTCGGATAACGCTGGTGGGAATATCGTTTACACAGACGCGTACGTTTGA 339897
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
DB 339898 GTAATCGACATAACTATTATTAACACCAACCAAGAAATACGTTATGGATATGCAACA 339957
QY 135 SerAsnAsnProAsnIleLysGlyGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
DB 339958 GATAAAATGGCAACGCTTTTAAACGGAACTTGCACCAATTCAAATCCCAATTAATAGC 340017
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
DB 340018 GCAATTCAGCGGAGCAACAAAAATGGTAGCTTAAGCGGTAACTTCCACTTGATTGG 340077

QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
DB 340078 GCGCAAAAA---GATACGATTTCTTCGAG--- 340104
QY 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
DB 340105 -----CTT 340107
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
DB 340108 TCTGTATACGATAATGCTGTGGAAACATAAATCAATCAATATGAAGCTGGACA 340167
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
DB 340168 CCAGATCGAGTGGCAATGTTCTCTATGAATATGAATCAATG---GAC 340215
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
DB 340216 GGAAGCGTTTAACTCTCCA-----GTAACAGGAACACATTAATAC 340257
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 340258 AACCGCAAGCGCACTTACAAACCCCTGATGCA--- 340290
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 340291 ---CTTAAACATTCATCAATTCACCTGTTAAACGCG--- 340326
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
DB 340327 -----AAACAAGCAAT 340338
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 340339 ATG-----GGCTTAAACCTAAGTGGCTTAAACA 340365
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
DB 340366 AACTAC-----GGTACCAACCAA 340383
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
DB 340384 GTATTCTCAACACTTCTGACGAAAGCGCTGCAACTGTAAAGACTACGCGATTACC 340443
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
DB 340444 GATTCTGTTATATTGCGAGTGAGTTACTCAGACGGTACAGTTATCCAGTTGCCCACTT 340503
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
DB 340504 GCGGTTCTACTTCTCAATGAAGACGCTTAGTCAAAATGGGGAATGGCAATATGTT 340563
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
DB 340564 CCAGGATTATCTTCTGGAGATGCGATTACGGGCTTGTGGCCAAACGAGCTGGCGA 340623
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
DB 340624 ATTAGCGGCTCTTCTATTAGAAGCTCAACGTTAGATTTATCCGCGAATTCGTTAACTTA 340683
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
DB 340684 ATGACATACCAAGTGGTTTCCAGGCAATACAAAGTTATCCGTTCTCAGATGACGCG 340743
QY 494 LeuGlnLysAlaLeuGluLeu 500
DB 340744 ATGAACCAAAATGTGAACCTTG 340764
RESULT 37
AX417039
LOCUS
DEFINITION
ACCESSION
SEQUENCE
4030 from Patent WO0228891.
PAT 02-SEP-2002
LINEAR
DNA
349980 bp
AX417039
DEFINITION
SEQUENCE
4030 from Patent WO0228891.
PAT 02-SEP-2002
LINEAR
DNA
349980 bp
AX417039
DEFINITION
SEQUENCE
4030 from Patent WO0228891.
PAT 02-SEP-2002
LINEAR
DNA
349980 bp

AUTHORS Kunst, P. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4033 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES Location/Qualifiers
Source 1. 349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"

/note="seq 2058, original length: 3,011,208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"

ORIGIN

Alignment Scores: 1.1e-19 Length: 349980
Pred. No.: 484.50 Matches: 137
Score: 42.21% Conservative: 77
Percent Similarity: 27.02% Mismatches: 188
Best Local Similarity: 18.74% Indels: 105
Query Match: 6 Gaps: 11
DB:

US-10-009-823A-1 (1-502) x AX417042 (1-349980)

QY 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 129803 ATGAATCAAACTATGATATACAGCTATTCTGGATGAATCGCTTCCAAACGATATACA 129862
QY 21 ThrValSerAsnAlaLeuAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 129863 GTGACATCAAAATATATGTCATGATGCAACACGACGAGGTATATAAAAACAGCGTCTT 129922
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db 129923 TTCAATGATTACTTTTACCAAAAACACATGGGATCTGTTGCAGCGCGACTTTATGCTGGA 129982
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 129983 ACAAAACCAATGAGCTCGGTTCCGGTTCGAAATGGGCGCAATTTAACCGATTATACA 130042
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
Db 130043 GCAGGTCTCCGACATCACTGCGCAAGAACAAAGATGCAGCACTCAAGCGCGCGCTTT 130102
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
Db 130103 TTCATTGCTGGCGAATACGCTGGTGGGAATATCGTTTACACGACGACGAGTCTTGGCA 130162
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db 130163 GTATCTGACATCACTATTATACACCGCAACAGGAAATACGTTATGGATATGCAACA 130222
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 130223 GATAAAATGCAACGTTTAAACCGAAATCTGCAACCAATTCAAATCCCAATTAATAGC 130282
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 130283 GCAATCCAGCGGAGCAACAAAATGGATGTAATGCGGTAACTCCACTTGATGG 130342
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSer 194
Db 130343 GCGCAAAAA--GATACGATTTCTTCGAG----- 130369
QY 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 130370 -----CTT 130372

QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
Db 130373 TCTGTATACATATCTCTGTGGAACATAAACTTCAAGTCAATATGAAGCTGGACA 130432
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 130433 CCAGATCGAGTGGCAATGTTCTCTATGATATGAATATCAAAATG-----GAC 130480
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 130481 GGAAACGGTAACTCTCTCA-----GTAACAGGAACACATTAAATTAC 130522
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 130523 AACCGCAAGCGCAACTTACAAACCCCTGATGCA----- 130555
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 130556 ---CTTAAACATTCAAATCAATCCACTGTTAAACGCG----- 130591
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 130592 -----AAACAGTCAAT 130603
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 130604 ATG-----GCTTAAACCTTAAGTGGCTTAACA 130630
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 130631 AACTAC-----GGTACCACCA 130648
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 130649 GTATTTCTCACCAACTTCTGACGAAAGGCGCTCAACTGTAAAGAACTAGCGAGTTACC 130708
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 130709 GATTCTGTTATATGCTGAGTACTCAGACGCTACAGTTATCCAGTTGCCCACTT 130768
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 130769 CGGGTGTCTACTTCTCCATGAAGCGCTTAGTCAAAATGGGGAATGCGCAATATGTT 130828
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 130829 CCAGGATTTATCTTGGAGATGCGAGTTTACGCGCTTGTGCGCCAAACGAGCTGGCGGA 130888
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 130889 ATTACGCGCTCTTCTATTAGAGGCTCAAACGCTAGATTATCCCGCAATTCGTTAACTTA 130948
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 130949 ATGACATACCAAGTGGTTTCCAAAGGCAATCAAAAGTTATCCGTTGTCAGATGACGTG 131008
QY 494 LeuGlnLysAlaLeuGluLeu 500
Db 131009 ATGAAACAAATGTGAAGTTG 131029
RESULT 39
AX641666 349980 bp DNA linear PAT 21-FEB-2003
LOCUS Sequence 2856 from Patent WO0101118.
DEFINITION AX641666
ACCESSION AX641666
VERSION AX641666.1 GI:28474427
KEYWORDS Listeria monocytogenes
SOURCE Listeria monocytogenes
ORGANISM Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1 Buchrieser, C., Frangeul, L., Couve, E., Rusniok, C., Fsihi, H.,
AUTHORS

Dehoux, P., Dussurget, O., Chetouani, F., Nedjari, H., Glaeser, P., Kunst, F., Cossart, P., Daniels, J., Goebel, W., Kreft, J., Kuhn, M., Ng, E., Vasquez-Boland, J., Dominguez-Bernal, G., Garrido-Garcia, P., Tierrez-Martinez, A., Amend, A., Chakraborty, T., Domann, E., Hain, T., Berche, P., Charbit, A., Durant, L., Perez-Diaz, J. C., Baquero, F., Garcia del Portillo, F., Gomez-Lopez, N., Maduenio, E., de Pablo, B., Weiland, J., Kaerst, U., Eniian, K. D., Hauf, J., Rose, M., and Voss, H.
 Listeria monocytogenes genome, polypeptides and uses
 Patent: WO 010118-A 2856 11-APR-2001;
 INSTITUT PASTEUR (FR)

TITLE JOURNAL

FEATURES

source

Location/Qualifiers

1. 349980
 /organism="Listeria monocytogenes"
 /mol_type="unassigned DNA"
 /db_xref="taxon:1639"
 /note="Seq 1 for big: original length: 2944528-splitted
 into 10 <23>seq 0001 : From 0.000.001 to 0.349.980
 length: 349.980 <23>seq 2855 : From 0.300.001 to
 0.649.980 length: 349.980 <23>seq 2856 : From 0.600.001
 to 0.949.980 length: 349.980 <23>seq 2857 : From
 0.900.001 to 1.249.980 length: 349.980 <23>seq 2858 :
 From 1.200.001 to 1.549.980 length: 349.980 <23>seq 2859
 2860 : From 1.500.001 to 1.849.980 length: 349.980 <23>seq
 <23>seq 2861 : From 2.100.001 to 2.449.980 length:
 349.980 <23>seq 2862 : From 2.400.001 to 2.749.980
 length: 349.980 <23>seq 2863 : From 2.700.001 to
 2.944.528 length: 244.528"

ORIGIN

Alignment Scores:

Pred. No.: 1.1e-19 Length: 349980
 Score: 484.50 Matches: 137
 Percent Similarity: 42.21% Conservative: 77
 Best Local Similarity: 27.02% Mismatches: 188
 Query Match: 18.74% Indels: 105
 DB: 6 Gaps: 11

US-10-009-823a-1 (1-502) x AX641666 (1-349980)

Qy 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 130516 ATGAATCAAACTAGTATACAGCTATTCTCGGATGAATCGGTCCCAACAGCATTATCA 130575
 Qy 21 ThrValSerAsnAniLeaAlaAsnAlaAsnThrIleGlyTyrllysGlnGlnValVal 40
 Db 130576 GTAACATCAAAATATATTCGCAATGCCAACACAGCAGGATATAAAAAACAACGCGTGT 130635
 Qy 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
 Db 130636 TTCAATGATTTACTTTTACCAAAAAACAATGGGATCTGTGACGGGACTTTATGCTGGA 130695
 Qy 57 GlnGlyProAsnGlnAlaGlyMetGlyValaGlnValGlySerValArgThrIlePheThr 76
 Db 130696 ACAAAACCAATAGTTTCGGTTCGGTTCGAAATTTGGGCGATTTAACTGACTATACA 130755
 Qy 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
 Db 130756 GCAGGTTCCCGGACTTCACTGGCAGAAATPAAAGATGCAGCAGCTACAAGCGCGCGCTTT 130815
 Qy 97 PheGlnVal-----ThrLeuGluAspLysValHisTyThrArgAlaGlyAsnPheArg 114
 Db 130816 TTCACTCTGGTGATAACGCTGTGGGATATCGTTTATACAGTCACGAGTACGTTTGA 130875
 Qy 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
 Db 130876 GTATCTGACAACTAATATTTAAACGCGCAGCAAGGAAATACGTTATGGGTTACGCAACG 130935
 Qy 135 SerAsnAspProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
 Db 130936 GACAAAATGGTACGTTTTAAACGGTAATTTTAAACCCATCAAAATCCCGTAAATAGC 130995
 Qy 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174

Db 130996 GCAATTCAGGGGAGCAACAAAAATGTTAGCTTAAGCGGTAAACATCCCACTTGATTGG 131055
 Qy 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyPheAlaLeuLeuGluSer 194
 Db 131056 GCGCAAAAA---GATACGATTCTTCCGAG----- 131082
 Qy 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTySerTyAlaGlnProMet 214
 Db 131083 -----CTG 131085
 Qy 215 ArgValTyAspGlnGlnGlyAsnSerHisAspIleThrValTyPheAspGlyAla--- 233
 Db 131086 TCTGTATACGATATGCTGGCGGAAACATATAACTTCAGTCAATATGAAGCAGCTACA 131145
 Qy 234 ProSerSerThrGlySerLysThrPheGluTyrlLeuValAlaMetAsnProSerGluAsp 253
 Db 131146 CCAGATCGCAGCGGTAAATGTTTCTCAGATATGAATTCAAATG-----GAC 131193
 Qy 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 Db 131194 GGCAAAGCATTAATCTCTCCA-----GTGACAGGACACTTAATTATC 131235
 Qy 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 Db 131236 AATGCGCAGGCGTAACCAACGACCCAGCGCA----- 131268
 Qy 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 Db 131269 ---CTTAAAAATATTCAAATCAATTCACAGTAACCGC----- 131304
 Qy 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 Db 131305 -----AAACAAGTCAAT 131316
 Qy 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 Db 131317 ATG-----GGCTTAAACCTTAAGTGGCTTAACA 131343
 Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
 Db 131344 AACTAC-----GGACGACCA 131361
 Qy 374 ThrArgGlySerGlnAspGlyTyProGlnGlyAspLeuValAspValThrIleThr 393
 Db 131362 GTATTCTCACCACACTCTGATGGCAAGGCTCTCGCACTGTAAAAGATTATGCACTTACC 131421
 Qy 394 SerGluGlyLysLeuGlnGlyLysTySerAsnSerGlnValValAspPheTyAsnIle 413
 Db 131422 GATTCTGCTATATTCGAGTGAGTATTTCAGATGGTACGATTATCCAGTTGCCCAACTC 131481
 Qy 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTySer 433
 Db 131482 CGCGTGTCTAGCTTCTCCATGAAGACGGTTAGTCAAAATGGGAAACGCGAATATGTT 131541
 Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyGlyLys 453
 Db 131542 CCAGGATATCTCTGCTGATGATGATGCGGCTCTGCGCAAAATGCGCTGCGGGA 131601
 Qy 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 Db 131602 ATTAGTGGTCTTCCCTAGAGGATCAACGAGACTTGTCCGCTGAATTCGTTAACTTA 131661
 Qy 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
 Db 131662 ATGACATATCAAGTGGTTTCCCAAGGGAATACAAAGGTTATTCTGTGTGGCGGATGACGTG 131721
 Qy 494 LeuGlnLysAlaLeuGluLeu 500
 Db 131722 ATGAACAAATTTGCACTTG 131742

RESULT 40
 AE015135

LOCUS AB015135 13374 bp DNA linear BCT 18-OCT-2002
DEFINITION Shigella flexneri 2a str. 301 section 98 of 412 of the complete
Genome.
ACCESSION AB015135 AB0005674
VERSION GI:24051365
KEYWORDS
SOURCE Shigella flexneri 2a str. 301
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 13374)
AUTHORS Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H.,
Liu, H., Yang, F., Qu, D., Zhang, X. B., Zhang, J. Y., Zhang, G. W.,
Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P.,
Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L.,
Qiang, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J.
TITLE Genome sequence of Shigella flexneri 2a: insights into
pathogenicity through comparison with genomes of Escherichia coli
K12 and O157
Nucleic Acids Res. 30 (20), 4432-4441 (2002)
JOURNAL 12384590
PUBMED 2 (bases 1 to 13374)
REFERENCE Jin, Q., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B.,
Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y.,
Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y.,
Lu, W. C., Qiang, B. Q., Wen, Y. M. and Hou, Y. D.
Direct Submission
TITLE Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
JOURNAL of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
China
FEATURES
source
1. 13374
/organism="Shigella flexneri 2a str. 301"
/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
111..1685
/gene="mw1n"
/locus_tag="SF1075"
111..1685
/gene="mw1n"
/locus_tag="SF1075"
/note="Residues 14 to 524 of 524 are 99 pct identical to
residues 1 to 511 of a 511 aa protein from Escherichia
coli O157:H7 ref: NP_309474.1"
/codon_start=1
/transl_table=1
/product="putative virulence factor"
/protein_id="AA042697.1"
/db_xref="GI:24051366"
/translation="MQAFCEWVIRTPMNLKSLAAVSSMTWFSRYLGFARDAIVARI
FGMATDAPFAFKLPNLRIPFAEGAFSAQFVPIABYKSKQGEDATRVFVYVSG
LITLALAVTVAGMLAAPWIMVTAPGFADTAPALTPSOLLKITPPYILLISLASLV
GAILNTWRESIPAFAPTLINISMGFALPAEPNPVPLALAWAVTVGIIQLVYQL
PHUKKGLMVLPRINFDAGMRVQKMPAILGVSVQISLIIINTIFASFLASGVSV
WMTYADLMFPFSGVLGVAUGTILLPSLKSFGASGNHDEYNRLMDGLRCFLALPS
AVALGILSGPTVSLFYQKFTADALMTQRIAYSVGLIGLIVVKVLPAGFYRQD
IKTPVKIAIVTILITQMLNLAFIGPLKHAGLSLSIGLAACNASLLYWLQRKQIFTP
QPGWMAFLRLVAVLVMGSVLLGMLHIMPENSLGTMPWRLRLMAVVLAGIAVFAA
LAVLGFVKFEFARTV"
complement (1725..2141)
/gene="flgN"
/locus_tag="SF1076"
complement (1725..2141)
/gene="flgN"
/locus_tag="SF1076"
/note="Residues 1 to 138 of 138 are 97 pct identical to
residues 1 to 138 of a 138 aa protein from Escherichia
coli K12 ref: NP_415588.1"
/codon_start=1
/transl_table=1
/product="protein of flagellar biosynthesis"

gene
CDS
/protein_id="AA042698.1"
/db_xref="GI:24051367"
/translation="MRLARILDMQSAVNLNDLKTWMDQEOHLSMGQINGSLQWITE
QKSSLLATDYLEQRKRPENTANSVDISQRWQEIIVTKTQQLRQMQRHNGWLLGGQIE
RNQQAEMLRKPHQEPFLYGANGQTSFPHRSGKKISI"
complement (2146..2439)
/gene="flgM"
/locus_tag="SF1077"
complement (2146..2439)
/gene="flgM"
/locus_tag="SF1077"
/note="Residues 1 to 97 of 97 are 100 pct identical to
residues 1 to 97 of a 97 aa protein from Escherichia coli
K12 ref: NP_415589.1"
/codon_start=1
/transl_table=1
/product="anti-FlaA (anti-sigma) factor; also known as
RifB protein"
/protein_id="AA042699.1"
/db_xref="GI:24051368"
/translation="MSIDRTSPKPVSTVQPRETTDAPVTNSRAAKTASTSTSVTL
DAQAKMQPGSSDINLRLVEALKLAINRNGELKMDTKIADALINEAQOQLQSN"
complement (2515..3174)
/gene="flgA"
/locus_tag="SF1078"
complement (2515..3174)
/gene="flgA"
/locus_tag="SF1078"
/note="Residues 1 to 219 of 219 are 98 pct identical to
residues 1 to 219 of a 219 aa protein from Escherichia
coli K12 ref: NP_415590.1"
/codon_start=1
/transl_table=1
/product="flagellar biosynthesis; assembly of basal-body
periplasmic P ring"
/protein_id="AA042700.1"
/db_xref="GI:24051369"
/translation="MLTIKRSVAIIAILFSPILSTASNLTSQHNFPQAQAGVSDVEV
VSIPTAPNLLPPCEQLLSMSNRMLNMGVNLARCGNDKRYLYQVYQATGNYVVAAM
PIVRGKLEAGNVKLRGLDTPPTVLDINQLVDVSLRDLSPDQFIQTQPRQAW
RVKAGQRVNVIASGDGFSANAEGQALNNAVAQNAVRYMVSEQVSVGVVDAGNLLIN
L"
3329..3745
/gene="flgB"
/locus_tag="SF1079"
3329..3745
/gene="flgB"
/locus_tag="SF1079"
/note="Residues 1 to 138 of 138 are 100 pct identical to
residues 1 to 138 of a 138 aa protein from Escherichia
coli K12 ref: NP_415591.1"
/codon_start=1
/transl_table=1
/product="flagellar biosynthesis, cell-proximal portion of
basal-body rod"
/protein_id="AA042701.1"
/db_xref="GI:24051370"
/translation="MLDKLDRLRFPQEQENLRQROEVLAAANIANADTPGVQARDID
FASLKKVMQGRDRTSVVLTMTSTQHPQAQLTPTAEQLYRIPDPSLDGNTVDM
DRERTQFDNLSQYQMSLSALSGQIKGMMNVLQSGN"
4164..6092
/gene="flgE"
/locus_tag="SF1080"
4164..6092
/gene="flgE"
/locus_tag="SF1080"
/note="Residues 241 to 642 of 642 are 98 pct identical to
residues 1 to 402 of a 402 aa protein from Escherichia
coli K12 ref: NP_415594.1"
/codon_start=1
/transl_table=1
/product="flagellar biosynthesis, hook protein"
/protein_id="AA042702.1"

Mon Nov 1 12:08:07 2004

```

Db      5802 CAAACCAACTGCTGGGCGAGATTGTACTGGCGAACTTTGCCAACACGAGGTCTGGCA 5861
QY      426 ArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu 445
Db      5862 TCCGAAGGCGACAACTGCTGGCTGGACGCAATCTTCTGGCGTGGCGCTGTGGGACA 5921
QY      446 ProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAsp 465
Db      5922 GCCGGACGGGAACTTTGGCACTCTGACCAACGGTGGCGTGGAAAGCGTCCAAACGTCGAT 5981
QY      466 MetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLys 485
Db      5982 CTCAGTAAAGAACTGTCAATATGATCGTTGCCACGCTAACTATAAGTCTAACGCCCCAG 6041
QY      486 SerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db      6042 ACCATCAAAACCCAGGACCATCCTCAACACGCGGGTTAACTTACGC 6089

```

Search completed: November 1, 2004, 06:05:13
 Job time : 6762 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:04:16 ; Search time 24 Seconds
(without alignments)
1387.151 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MWGSLFIGATGKMKTHSTGLG.....NSKSVTTADTWMQKALELKR 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 segs, 56318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/aaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/1/aaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/1/aaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/1/aaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/1/aaa/ECTUS_COMB.pep:*
- 6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	528.5	20.4	471	4	US-09-252-991A-19526
2	475.5	18.4	412	4	US-09-543-681A-8182
3	313	12.1	265	3	US-08-483-857-8
4	307	11.9	269	4	US-09-252-991A-19715
5	287	11.1	261	3	US-08-483-857-10
6	286	11.1	265	4	US-09-543-681A-8110
7	285	11.0	264	2	US-08-436-748-8
8	237.5	9.2	263	3	US-08-483-857-5
9	229.5	8.9	262	2	US-08-436-748-7
10	228.5	8.8	263	3	US-08-483-857-7
11	223.5	8.6	263	2	US-08-436-748-5
12	220	8.5	260	2	US-08-436-748-10
13	176	6.8	2736	4	US-09-252-991A-30227
14	173.5	6.7	1095	4	US-09-206-942-69
15	173.5	6.7	1536	1	US-08-038-682-2
16	173.5	6.7	1536	1	US-08-302-832-2
17	173.5	6.7	1536	2	US-08-530-198-2
18	173.5	6.7	1536	2	US-08-469-880-2
19	173.5	6.7	1536	2	US-08-728-470-2
20	173.5	6.7	1536	2	US-08-617-697-2
21	173.5	6.7	1536	3	US-08-719-641-2
22	173.5	6.7	1536	4	US-09-206-942-67
23	169.5	6.6	248	2	US-08-436-748-6
24	165	6.4	247	3	US-08-483-857-6
25	155	6.0	270	2	US-08-436-748-3
26	155	6.0	270	3	US-08-483-857-4
27	149.5	5.8	1160	4	US-09-206-942-65

28	149.5	5.8	1188	4	US-09-206-942-63	Sequence 63, Appl
29	143	5.5	1306	4	US-09-538-092-330	Sequence 330, App
30	142	5.5	1095	4	US-09-206-942-45	Sequence 45, Appl
31	142	5.5	1101	4	US-09-206-942-43	Sequence 43, Appl
32	140.5	5.4	529	4	US-09-381-656-1	Sequence 1, Appl
33	137	5.3	585	4	US-09-655-479A-20	Sequence 20, Appl
34	136.5	5.3	2383	4	US-09-622-709A-302	Sequence 302, App
35	133	5.1	1833	4	US-08-621-944A-4	Sequence 4, Appl
36	133	5.1	1833	4	US-08-945-567D-4	Sequence 4, Appl
37	133	5.1	1992	4	US-08-621-944A-3	Sequence 3, Appl
38	133	5.1	1992	4	US-08-945-567D-3	Sequence 10, Appl
39	132	5.1	2123	3	US-08-968-685A-10	Sequence 49, Appl
40	132	5.1	2314	3	US-09-268-347-49	Sequence 48, Appl
41	131	5.1	2048	3	US-09-336-115C-4	Sequence 4, Appl
42	130	5.0	733	4	US-09-538-092-1160	Sequence 1160, Ap
43	129.5	5.0	1475	4	US-09-538-092-1081	Sequence 1081, Ap
44	128	4.9	2090	4	US-09-540-236-3459	Sequence 3459, Ap
45	127.5	4.9	2142	4		

ALIGNMENTS

RESULT 1
US-09-252-991A-19526
; Sequence 19526, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19526
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19526

Query Match 20.4%; Score 528.5; DB 4; Length 471;
Best Local Similarity 28.5%; Pred. No. 8.4e-39;
Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;

QY	4	SLFTGATGKMKTHSTGLGTVSNINIANANTI	GYKQQQVVFQDLFSQDLAIGTSGSQGPQAG	63
Db	11	SNFNLGSLGIAQAASSGLNVGTGNNIAGTVGFKQSR	AEAFADYAAASV-LGS-----GSNPQG	65
QY	64	MGAVGVSRTIFTGAFEPGNSVTDLAIGKGFQVTL	EDKVVHVRAGNREFTDGFLND	123
Db	66	SGVLSDVSQMFQKGNIDSTNSVLIDLAINGPFVT	NNGAISYTRAGYFNTDKQFIVD	125
QY	124	PSGFTLMGSRISNNPNIKKETLEPIQLD-FNDPT	VAKSPARTSTALNAVNVLGSDTDKTK	182
Db	126	NNGYRLQGYAVGPNQGLQNGVWTDLKV	ERANQAQATSSIQQSYNLNSLTK-----P	177
QY	183	SEAMPYFALLESWKNGTTPPISTNSY	YAQPMRYVDQGNSHDITVYF---DGAPST--	237
Db	178	FTVTFP-----DPSDAATYNSSSSLGYDSQ	GNSHNTMSQFFIKNEPDPNATPP	225
QY	238	---GSKTFEYLV-AMNPS	EDCSAASGTDLSAGLMSGTMTFSSNGELKNM	290
Db	226	IPENSWTKVLIDGVNPLDPSNKT-----MS	FNVTFDASQMTSVRA--PDGSTG	275
QY	291	-----ATKDLN	AWQAPLVNGLPQFSANFVGAGIQPLTLD	344
Db	276	PGFSIDATNTVIQFSPA---TGNEPTP---	GTGWIPASD-GKTPPTYAMNGATGAASG	327

QY 345 IGTDIGKLPSPMPITQSSGNSTARNSSSTRYS-----QDGYPOGDLVDVITSE 395
Db 328 ISFVRK-----TQISTAFQSNPIQDGYTGGLAGLEIDT 365
QY 396 GKQGGKYSQVDFVYNIPIARFTSEGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLS 455
Db 366 GVIFARYTNGSKVQGVWLANFANIQGLTPIKTSWQSAESGEPAVGAPRSGTLGALQ 425
QY 456 VNOLETNSVDMSEVMNMIIOGFOMNSKSVTTADTLMQKALELK 501
Db 426 SGALESNVDSINELVNLVHQRYQANAKTIQTEDAVTITINLR 471

RESULT 2

US-09-543-681A-8182
; Sequence 8182, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8182
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8182

Query Match 18.4%; Score 475.5; DB 4; Length 412;
Best Local Similarity 26.7%; Pred. No. 3.9e-34;
Matches 136; Conservative 81; Mismatches 165; Indels 127; Gaps 15;
QY 10 TGMKTHSTGLTGSNNIANANTIGYKQOVVFDLPSQDLAIGTSGSQPNQAGMAQVG 69
Db 14 SGLNAAANLDTGNINSATYGFKGATVSFADVFA-----GS-----GAGLGVRYS 61
QY 70 SVRTFTQGAPEPCNSVTDLAIGKGFQVTLDEK---VHYTRAGNFRFTQDGLNDPSG 126
Db 62 GISQNFQDGSITITNRTDVAISGGFFRI--EDSNGGVFYRNGEFGKXGKYLNNMG 119
QY 127 FTLGMSRISN---NPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVNLG---DSTD 179
Db 120 MRITGYPVQVVDGKNVYQKGTATPTII---PTDMYN-ASATDKMDMTVNLNSAEAD 174
QY 180 KTOSEANPYFALLESWKNGTTPISTSNYSYAQPMRVYDQGNHSDITVYFDGAPSGS 239
Db 175 QYTHKDP-----KDN-----DSYFNSTNVTYDLSGNEHNLNLP-----VKT 213
QY 240 KTEFYLVAAMPBDSGAASGTSAGLLMGTMTFSSNG-----ELKNMTAFTPTGSAT 292
Db 214 KDNWSVYAQDTTGTGPAQDL-----GKLYKONGVLDETAPKLNFTTVAYKGS--- 263
QY 293 KDLNAMQAPLVNGLPQFSANFVGAGIQPLTDLFGIKSQQNMWAGAPASAAIGTDIGKL 352
Db 264 -----QPMDEMNF----- 272
QY 353 PSMMPITQSSGNSTARNSSSTRYSQDGYPOGDLVDVITITSEGLQGYKYSNVQVDFYN 412
Db 273 -----SGSTQKVAESSVSKLAQNGYQAGEFTNRIEIPDGSIMATYSNQSQSVQV 323
QY 413 IPLARFTSEGLRREGNNHYSATLDSGGPEFGLPGTSNYKLSVNOLETNSVDMSEVMN 472
Db 324 IALANFANPGLSSQGNMMWSETNGSGSPVGVAGSGVFKLNTNNALEASNVDMSELVN 383
QY 473 MIIOGFOMNSKSVTTADTLMQKALELK 501
Db 384 MIVAQRYQSNQACTIKTQDQILQTLVSMR 412

RESULT 3

US-08-483-857-8
; Sequence 8, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-483-857-8

Query Match 12.1%; Score 313; DB 3; Length 265;
Best Local Similarity 21.9%; Pred. No. 7.2e-20;
Matches 111; Conservative 46; Mismatches 101; Indels 250; Gaps 11;
QY 1 MMSLFTGATGKMTHTSTGLTGSNNIANANTIGYKQOVVFDLPSQDLAIGST-----G 55
Db 1 MLRLSYGISGGKRFQKLDVIGNNIANVAVTGFKSRVTFKDMVSTQIAGSAAAGATIG 60
QY 56 SQGNQAGMAQVGSVRTFTQGAPEPCNSVTDLAIGKGFQVTLDEKVKHYTRAGNFR 115
Db 61 GTNSKQJGLSSSGTIDTHSTATQSTGTDLDAIDGCGYFRIDTCDGTAYTRAGNF-- 118
QY 116 TQDGLNDPSGFTLMGSRISNPNIKETLEPIQLDNDPTVAKSPAKTSTALNAVNLG 175
Db 119 ----- 118
QY 176 DSTDKTQSEANPYFALLESWKNGTTPISTSNYSYAQPMRVYDQGNHSDITVYFDGAPS 235
Db 119 ----- 118
QY 236 STGSKTEFYLVAAMPBDSGAASGTSAGLLMGTMTFSSNGELKNMTAFTPTGSATKDL 295
Db 119 -----DNTGLV-----TGDSYHVL 135
QY 296 NAWQAPLVNGLPQFSANFVGAGIQPLTDLFGIKSQQNMWAGAPASAAIGTDIGKLPSM 355
Db 136 N-----MNG-----GTIKPTD-----AQSFSLGSD-----SK 158
QY 356 MPIQTSSGNSTARNSSSTRYSQDGYPOGDLVDVITITSEGLQGYKYSNVQVDFYNIP 415
Db 159 VSI-----VDAEGKTQ-----DGGQIGI 176

QY 416 ARFTSGDGLRRGNHYSATLDSG-GPEFGLPGTSGNYGKLSVNOLETNSVDMREVMNMI 474
Db 177 VTFANSDDLKIGSNLYRESLNSGTASAAQDGGTGAIKSGFLENENVDLTDEFTMI 236
QY 475 IIRGFQFMSKSVTTADTWMQKALELKR 502
Db 237 VAQRGFQNSKIITTSDEILOBLVNLKR 264

RESULT 4

US-09-252-991A-19715
; Sequence 19715, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 19715
LENGTH: 269
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19715

Query Match 11.9%; Score 307; DB 4; Length 269;
Best Local Similarity 21.6%; Pred. No. 2.5e-19; Indels 254; Gaps 11;
Matches 108; Conservative 46; Mismatches 93

QY 1 MMGSLFIGATGKTHSTGLGTGVNNIANANTIGYKQQQVVFQDLFSQDLAIGSTGSGQPN 60
Db 9 MSLALVWSTGLSAQDMNLTISNLANVSTGFKEDRAEFQDLVQIRRRQGGQSTQDS 68
QY 61 QAGMGAVQS-VRTTITQAGPFGNSVT-----DLAIGKGFFOVTLED-KVHYTRAGNF 113
Db 69 ELPSGLQLGTGRVVGVTQKIFTPGSLQTTEQPLDMAVNGRGPVLLPDGTGVSYTRDGSF 128
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAV 173
Db 129 HLNSDQIVTNGFAL-----EPAI----- 148
QY 174 LGDSTDKTOSEANPYFALLESMKNGTTPPSTSNYSYAOPMRVYDQGNSHDITVFDGA 233
Db 149 ----- 148
QY 234 PSSTGSKTFEYLVAAMPSPEDGSAASCTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSA 293
Db 149 -----VVPE-----TQFT----- 158
QY 294 DLNAPAPLVNGLPOFSANFVGAGIQPLTLDPGIKSQONMWAGAPASAAAIGTDIGKLP 353
Db 159 -----VG----- 160
QY 354 SMPIQTSSGNSTARNSSSTRYSQDGYPGQDLVDVTITSEGLKQKYSNQVDFYNI 413
Db 161 -----QDG-----TVSVTTTNGAQ-----POVIG--NI 181
QY 414 PLARFTSEDGLRREGNNHYSATLDSGPEFGLPGTSGNYGKLSVNOLETNSVDMREVMN 473
Db 182 QTADFPINAGLQAIIGNNLFLETGSSGAPQVGTGGLNGLGTVAQNTLNSNVNVEELVNM 241
QY 474 IIRGFQFMSKSVTTADTWM 494
Db 242 IITQRAYEMNSKVIISTADQML 262

RESULT 5

US-08-483-857-10
; Sequence 10, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:

APPLICANT: Chan, Voon Loong
APPLICANT: Louie, Helena
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
TITLE OF INVENTION: CAMPYLOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,857
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-857-10

Query Match 11.1%; Score 287; DB 3; Length 261;
Best Local Similarity 21.2%; Pred. No. 1.5e-17; Indels 258; Gaps 10;
Matches 108; Conservative 42; Mismatches 101

QY 1 MMGSLFIGATGKTHSTGLGTGVNNIANANTIGYKQQQVVFQDLFSQDLAIGSTGSGQPN 60
Db 1 MISSLIWIAKTGLDQAQTMNDVIANLANVSTNGTKRQRAVFDLLYQ--TIROPGAQSS 58
QY 61 QAGM-GAVQS-VRTTIT-----OGAFEPGNSVTDLAIGKGFFOVTLEDKVH-YTRAG 111
Db 59 QTTLPGLQIGTGRVPVATERLHSGNLSTNNKDVAIKQGGFFQVWLPDGTSAITRDG 118
QY 112 NFRFTQDGLNDPSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAV 171
Db 119 SFGVDGQQLVTAGGFQV-----PAITIPA----- 144
QY 172 VNLGSDTKTOSEANPYFALLESMKNGTTPPSTSNYSYAOPMRVYDQGNSHDITVFD 231
Db 145 -----NALSIITGRD 154
QY 232 GAPSTGSKTFEYLVAAMPSPEDGSAASCTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSA 291
Db 155 GWSVT-----QQQAA----- 166
QY 292 TKDLNAPAPLVNGLPOFSANFVGAGIQPLTLDPGIKSQONMWAGAPASAAAIGTDIGK 351
Db 167 ----- 166
QY 352 LPSMPTQTSSGNSTARNSSSTRYSQDGYPGQDLVDVTITSEGLKQKYSNQVDFY 411
Db 167 -----PVQVGQLN----- 174

QY 412 NIPLARFTSEDGLRREGNNHYSATLDSGGPEFGLPGTNSYKLSVNOLETNSVNDMSREMY 471
Db 175 ---JTFWNTDGLSEIGENLYETQSSGARNSTPGLNGAGLLYQGYVEITSNVNVAELV 231
QY 472 NMIIQRFQWNSKSVTTADTMQLKALEL 500
Db 232 NMIIQVQAYEINSKAVSTTDQMLKLTQL 260

RESULT 6

US-09-543-681A-8110
; Sequence 8110, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8110
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8110

Query Match 11.1%; Score 286; DB 4; Length 265;
Best Local Similarity 19.7%; Pred. No. 1.9e-17; Mismatches 55; Indels 254; Gaps 9;
Matches 100; Conservative 55; Mismatches 98; Indels 254; Gaps 9;
QY 1 MMSLFLGATGKMTHTGLGTGVSNNIANANTIGYKQOVVFDLFSQDL----AIGSTGS 56
Db 6 MIRSLWIAKTGLDAQQTMDVSNLNANVSTNGFKRQAVFEDLLYTIRQPCAMTSEQT 65
QY 57 QGPN--QAGMGAQVGSVRTFTQGAPEFGNSVTDLAIGKGFQVLTEDKVH-YRAGNF 113
Db 66 NAPSGLQIGTVGRFVATERLHSGNLAQNTGTRDVAIKQGFHFVQLPGDTAYTRDGSF 125
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNV 173
Db 126 QMDQNGOLVTSQGFQVPA-----ILPETAKKVV-----156
QY 174 LGSDTKTQSEANPYFALLESWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVFDGA 233
Db 157 ---PQ-----GRDGIVSVEIGS 169
QY 234 PSSTGTSTKTFEYLVAMPNSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293
Db 170 PA-----171
QY 294 DLNAPAPLVNGLPQFSANFVGAGIQPLTLDFGIKSOQNMWAGAPASAAAGTIDIGKLP 353
Db 172 ---PQ-----QVQQL- 178
QY 354 SYMPIQTSSGNSSTARNSSSTRYSQDGYFQGDVLDVTITSEKLGKYSNSQVDFYNI 413
Db 179 ---TLTT-----182
QY 414 PLARFTSEDGLRREGNNHYSATLDSGGPEFGLPGTNSYKLSVNOLETNSVNDMSREMYNM 473
Db 183 ---FINDSGLESVGENLYLETASSCAPTENAPGINGAGLLYQGYVETSNVNAEELVNM 238
QY 474 IITQRFQWNSKSVTTADTMQLKALEL 500
Db 239 IQTRAYEINSKAISTSDQMLKLTQL 265

RESULT 7

US-08-436-748-8
; Sequence 8, Application US/08436748

; Patent No. 5827654
; GENERAL INFORMATION:
; APPLICANT: CHAN, YOON LOONG
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,748
; FILING DATE: 05-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-428 MIS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-436-748-8
Query Match 11.0%; Score 285; DB 2; Length 264;
Best Local Similarity 20.3%; Pred. No. 2.3e-17;
Matches 103; Conservative 47; Mismatches 108; Indels 250; Gaps 8;
QY 1 MMSLFLGATGKMTHTGLGTGVSNNIANANTIGYKQOVVFDLFSQDLAIGSTGSQ-- 58
Db 1 MSLALYSGISGKNFQKLEVI GNNLANMSTVGTQRVTFKDMVSTIAGGSAAGGIG 60
QY 59 ---PQAGMGAQVGSVRTFTQGAPEFGNSVTDLAIGKGFQVLTEDKVHYTRAGNFRF 115
Db 61 GTNSKQIGLGGSSGTIDTIHSTSATQTTGRTKDLAIDGDMYRVDGSGEAYTRAGNF-- 118
QY 116 TQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNVNLG 175
Db 119 ---YLNNTGTLVTGGS-----YHVVNMN 138
QY 176 DSTDKTQSEANPYFALLESWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGAPS 235
Db 139 GGTIKIPTDAQSF-----151
QY 236 STGSKTEYLVAMPNSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATKDL 295
Db 152 -----151
QY 296 NAMOPAPLVNGLPQFSANFVGAGIQPLTLDFGIKSOQNMWAGAPASAAAGTIDIGKLP 355
Db 152 -----SIGSD-----SK 158
QY 356 MPIQTSSGNSSTARNSSSTRYSQDGYFQGDVLDVTITSEKLGKYSNSQVDFYNIPL 415
Db 159 VQI-----VDAQGKTQ-----DGGQIGI 176
QY 416 ARTPTSEDGLRREGNNHYSATLDSG-GPEFGLPGTNSYKLSVNOLETNSVNDMSREMYNM 474
Db 177 TTFANSQDLKIGSNLPRESLNSGTASEANQPADGATAALKSTYLENSVNDVDTDETEMI 236

Db 143 -----PQATDI 149
Qy 227 TVYFDGAPSTGKTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKNMTAFT 286
Db 150 TI-----SK-----153
Qy 287 PTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKSKQNNWAGAPASAAAIG 346
Db 154 -----SLVQV-----159
Qy 347 TDIGKLPMMPIOTSSGNSTARNSSSTRYSODGYPQGLVDVITITSEKLGKYSNSQ 406
Db 160 -----KL-----DGPQPV-----GOIQ-----174
Qy 407 VVDYFNIPLARFTSEDLRRREGNNHYSATLDSGGPEFGLPCTSYGKLSVNQLETSNVD 466
Db 175 -----LANFLNEGGLAIGDNLFLTAASGAATLVRASRALACCCSTDTTEASNVDA 226
Qy 467 SREVMNMIIRGFQMSKSVTTADTMLOKALEIK 501
Db 227 VSEITALITAGRAYEMNSKVISTADQMLQATSQLR 261

RESULT 10

US-08-483-857-7
; Sequence 7, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-483-857-7

Query Match

Best Local Similarity 8.8%; Score 228.5; DB 3; Length 263;
Matches 106; Conservative 32; Mismatches 109; Indels 269; Gaps 11;

Qy 2 MGSIFGATGKTHSTGLGVSNNNANNTIGYKQCVFQDLFSQDL-AIGSTGSQGN 60
Db 1 MQALRTAASGMAAQLNVEVISNNIANNMTVGFKRAAEFQDLLYTIERRAGSQSSTGDN 60

Qy 61 -----QAGMGAQVGSVRTIETQGAPEFGNSVTDLAIGGKGFQVTL-EDKVHYTRAGNF 113
Db 61 IVPTQVGGVKGAGSVYRITTEQGTPTLTDSPDLAIOGKGMFILLPSGGTATRAGNF 120
Qy 114 R-----FTQDGLNDPSPGFTLMGRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTST 166
Db 121 STNDQGIQVTEGGLVQV-GITP-----143
Qy 167 ALNAVNLGDSITQKTSQSEANPYFALLESWKNGTPISTSNYSYAQPMRVYDQGNSHDI 226
Db 144 -----143
Qy 227 TVYFDGAPSTGKTFEYLVAAMPSEDSAAAGTDSAGLLMSGTMTFSSNGELKNMTAFT 286
Db 144 -----QNAIT 150
Qy 287 PTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKSKQNNWAGAPASAAAIG 346
Db 151 ISKS-----GLVQV-----159
Qy 347 TDIGKLPMMPIOTSSGNSTARNSSSTRYSODGYPQGLVDVITITSEKLGKYSNSQ 406
Db 160 -----KL-----DGPQPV-----GOIQ-----174
Qy 407 VVDYFNIPLARFTSEDLRRREGNNHYSATLDSGGPEFGLPCTSYGKLSVNQLETSNVD 466
Db 175 -----LANFLNEGGLAIGDNLFLTAASGAATLVRASRALACCCSTDTTEASNVDA 226
Qy 467 SREVMNMIIRGFQMSKSVTTADTMLOKALEIK 502
Db 227 VSEITALITAGRAYEMNSKVISTADQMLQATSQLR 262

RESULT 11

US-08-436-748-5
; Sequence 5, Application US/08436748
; Patent No. 5827654
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: LOUIE, HELENA
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,748
; FILING DATE: 05-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-428 MIS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-436-748-5

Query Match 8.6%; Score 223.5; DB 2; Length 263;
Best Local Similarity 19.3%; Pred. No. 7.5e-12;
Matches 98; Conservative 43; Mismatches 115; Indels 253; Gaps 12;
QY 1 MMSLPIGATGKMTHTSTGLTVSNNIANTTGYKQQQVVDLFSQDLAIGSTGSGPN 60
DB 1 MMSLPIGATGKMTHTSTGLTVSNNIANTTGYKQQQVVDLFSQDLAIGSTGSGPN 60
QY 61 QAGMGAGVG-----SVRTFTQGAPE-PGNSVTDLAIGGKGFQVOTLEDKVHYTRAGNF 113
DB 61 LPSGLIEGVGVPRVAVTKVTFEGNLSKSTSTGLDWAAGNGFFQIQL-----108
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNKKTETLEPIQLDFNDPTVAKSPAKTSTALNAV 173
DB 109 -----PDCTIEMG-----NLQK-----IMK-----123
QY 174 LGDSTDKTQSEANPYFALLESKWNGTPTISTSNYSYAQPMRVYDQGNSHDITVYF 233
DB 124 -----DNEGNI-----129
QY 234 PSTGSKTPEYLVAMNPSEDGSAAGTDSAGLMSGTMTFSSNGELKNMTAFTPTGSA 293
DB 130 -----VNSDGYRLLPEMTI-----PEG-----146
QY 294 DLNMQAPLVNGLPOFSANFVAGIQLPLDPLFGIKSQQNMWAGAPASAAAIGTDIG 353
DB 147 -----ATAINVATD-GTVS 159
QY 354 SMVPIQTSSGNSTARNSSSTRYSQDGYPOGLVDVTITSEKLGKYSNQVVDVYNI 413
DB 160 VMLP-----GEQQTQI---GQV 174
QY 414 PLARFTSEDLRREGNNHYSATLDSGGPEGLPGTSNYGKLSVNQLETNSVDMBSREM 473
DB 175 ELVQFINPAGLHSMGDNLYLETGASGAPVAGIAGQDGLGTLIRHGFIELSNVQLEMTDL 234
QY 474 IIRQGFOMNSKSVTTADTMLOKALELKR 502
DB 235 ITQRAYEAGSKAITTSDMDLGIIVNQLKR 263

RESULT 12
US-08-436-748-10
Sequence 10, Application US/08436748
Patent No. 5827654
GENERAL INFORMATION:
APPLICANT: CHAN, YOON LOONG
APPLICANT: LOUIE, HELENA
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
TITLE OF INVENTION: CAMPYLOBACTER
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,748
FILING DATE: 05-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-428 MIS
TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 260 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-748-10
Query Match 8.5%; Score 220; DB 2; Length 260;
Best Local Similarity 19.2%; Pred. No. 1.5e-11;
Matches 98; Conservative 49; Mismatches 103; Indels 260; Gaps 12;
QY 1 MMSLPIGATGKMTHTSTGLTVSNNIANTTGYKQQQVVDLFSQDLAIGSTGSGPN 60
DB 1 MISALIAKSLDAAQQTNNMEVIANNLNMGNTGKRAVFEDLLYQ--TIROFGAQSSE 58
QY 61 QAGM--GAQVGS--VRTFT-----QGAPEPGNSVTDLAIGGKGFQVOTLEDKVHYTRA 110
DB 59 QGNITPGLQIGTVKAVATERLHSSQGLTQTNNS--KDAVKGGQYMPVMLPSGESAYTRD 117
QY 111 GNFRFTQDGLNDPSGFTLMGSRISNNPNKKTETLEPIQLDFNDPTVAKSPAKTSTALNA 170
DB 118 GSGVNGNGQLVTAGGFQV-----QFAITIPA---144
QY 171 VVNLGDSDTKQSEANPYFALLESKWNGTPTISTSNYSYAQPMRVYDQGNSHDITVYF 230
DB 145 -----NALSIHGR 153
QY 231 DGAPSTGSKTPEYLVAMNPSEDGSAAGTDSAGLMSGTMTFSSNGELKNMTAFTPTGS 290
DB 154 DGVVQVT-----QGGQAA-----166
QY 291 ATKDLNMQAPLVNGLPOFSANFVAGIQLPLDPLFGIKSQQNMWAGAPASAAAIGTDIG 350
DB 167 -----PVT-----VG 171
QY 351 KLPSWMPITQSSGNSTARNSSSTRYSQDGYPOGLVDVTITSEKLGKYSNQVVDV 410
DB 172 QL-----173
QY 411 YNIPARFTSEDLRREGNNHYSATLDSGGPEGLPGTSNYGKLSVNQLETNSVDMBSREM 470
DB 174 ---NLTTFMNDTGLSIGENLFIETQSSGAANESTPALNAAALLYQTVVTSNVNVA 230
QY 471 VNMIIQGFOMNSKSVTTADTMLOKALEL 500
DB 231 VNMIIQGFOMNSKSVTTADTMLOKALEL 260

RESULT 13
US-09-252-991A-30227
Sequence 30227, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30227
LENGTH: 2736
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match 6.8%; Score 176; DB 4; Length 2736;
Best Local Similarity 22.8%; Pred. No. 5.4e-06;
Matches 120; Conservative 56; Mismatches 212; Indels 142; Gaps 22;
QY 20 GTVSNNIANTIGYKQOQVVDLFSQDLAIGTSGSQG-----PNOAGMGAQGVSV 71
DB 1320 GTVVNAVA-----QDPA-GNTGPGQSTTVDAVAPFPVVPNSGNL 1359
QY 72 RTFTQGAPEPNSVTDLAIGKGFQVTLDEKHYTRAGNFRFTQDGFNDPSGFTLMG 131
DB 1360 ----LNGTAEPGSVTLTDGNGNPIGQTADG-----SGNWSFTPSQPLNGTVVNTA 1409
QY 132 SRISSNNPNKKELEPIQLDNDPTVAKSPAKTSTALNAVNLGSDTDKTSQSEANPYFAL 191
DB 1410 SDAAGNTSAPATT---TVDSLSLPIQVDPSPNGSVISGTADAGNTIIITDGNNGNPIGVQV 1465
QY 192 LESWKGNG--TPPISTSNYSVAQPMRVYDQGNSHDITVYFDG-APSSGSKTFEYLIVAM 248
DB 1466 TADSGGWSFTPGIPLDPGTVVNVARSNSVDSAPAVITVDGAPAAP-----VI 1516
QY 249 NPSEDGSAAGTDSAGLIM-----SGTMTFS-----SNGELKNMTAFT 286
DB 1517 DPS-NGTEISCTAAGATVILTDGNGNPIGQATADGSGNMTFTPGTPLANGTVINAVAQD 1575
QY 287 PTG-----SATKLNWQAPLIV-----GLPQFSANFV-----GAGIQPLTLDG 327
DB 1576 PAGNTSGPASVTDAAIA-PPAPVINPSNGVVISGTAAGATVILTDGNGNPIGQVTDAGS 1634
QY 328 IKSQONWAGAPASAAAGTIDIGKLPSPMPTQSSGNSSTARNSSSTRYSQDGPQDGL 387
DB 1635 GK-----WAFPTATPLANGTVINAL-----AQDAAGNNSPTSATVDSLAAPV 1679
QY 388 VDVITSEGLKQKYSQVDFNPIARFTSEDLRRREGNNHYSATLDSGPEFGLPG 447
DB 1680 IDP-----SNGSVIAGTAAGATVILTDG-----NGNPIGQVTDAGSGNWSFTFG 1724
QY 448 TSNGYKLSVNLSTSNVDMREVMNIIIQGFQNSKSVTTADTLMQKA 497
DB 1725 T-----PLSNGTV-----VNAVAQAAGNTSGPASTTVDSVAPAA 1759
RESULT 14
US-09-206-942-69
; Sequence 69, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-69
Query Match 6.7%; Score 173.5; DB 4; Length 1095;
Best Local Similarity 22.8%; Pred. No. 2.1e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKHTSTGLTVSNNIAN-TIGYKQOQVVDLFSQDLAIGTSGSQGNQAGMGAQGVSV 70
DB 570 IKAGVDGNSDSATNNANLTIKTKELK-----TQDLNI-----SGFNKAETAKDGS 618
QY 71 VRTFTQGAPEPNS-----VTDLAIGKGFQVTLDEKHYTRAGNFRFTQDGFND 123

DB 619 DLTIGNTSADGNAKKTFTQVKDSKISADG-HKVTILHSKVE--TSGSNNNTEDSSDN 675
QY 124 PSQFILMSGRISNNPNKKELEPIQLDNDPTVAKSPAKTSTALNAVNLGSDTDKTSQ 183
DB 676 -AGLTIDARNVNNNITSHKAVSI-----SATSGEITTKTGTINATGNVEITAOQGS 729
QY 184 EANPYFALLESKWNGT-----PPISTSNYSVAQPMRVYDQGNSHDITVYFDGAPSTG 238
DB 730 ----ILGIEISSGSVTLTATEGALAVNIS-----GNTVTVTAN-SGALTTLA 773
QY 239 SKTFEYLIVAMPSEDSAAAGTDSAGLL---MSGTMTFSSNGELKNMTAFTPTGSATKOL 295
DB 774 GSTIKGTESVTTSSQSGDIGGTISGTVVEKATESLITQSNKIKATTGEANVTSAITGI 833
QY 296 NAWQAPLIVNGLPQFSANFVGAGIQLTLDGFIKSQONWAGAPASAAAGTIDIGKLPSP 355
DB 834 GGTISGNTV-----VTANAGDLIVNGAE-----INATEGAATVITSSGKL--- 875
QY 356 MPQTSNGNSSTARNSSSTRYSQDGPQDGL--VDVITSEGLK---QGK----- 401
DB 876 ----TTEASSHITSAGQVNLQAQDGSVAGSINAANVTLTGTLTTVKGSNINATSGTL 931
QY 402 YSNQVDFNPIARFTSEDLRRREGNNHYSATLDSGPEFGLPGTSNYKLSVNOLET 461
DB 932 VINAKDAELNGAALGNHTVVNATNANGSGSVIATSS---RVNITG---DLITINGL-- 982
QY 462 SNVDMREVMNIIIQGFQNSKSVTTADTLMQKALELKR 502
DB 983 -NI-ISKNGINIVLL-KGVKIDVKYIQFGIASVDEVIEAKR 1020
RESULT 15
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```
US-08-038-682-2
Query Match      6.7%; Score 173.5; DB 1; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKTHSTGLGTVSNINAN-TIGYKQOVVFDLPSQDLAIGSTGSGPNQAGCAQVGS 70
Db 1011 IKAGVDGSDSDATNNANLTKTKELK-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTITQGAPEPNS-----VTDLAIGKGFFQVLTEDKVHYTRAGNFRFTQDGFND 123
Db 1060 DLTIGNTSADGTNAKVTFNQVKDSKISADG-HKVTLSKVE--TSGSNNTEDSSDN 1116
QY 124 PSQFTLMGSRISNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSDTKTQS 183
Db 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGETITTKTGTINATTGNVEITAGTGS 1170
QY 184 EANPYFALLESKWNGCT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSG 238
Db 1171 -----ILGIESSGSVTLTATEGALAVSNIS-----GNVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVAMNPSDEGSAAGTDSAGLL--MSGTMTFSSNGELKNMTAFTPTGSATKOL 295
Db 1215 GSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAMQAPLVNGLPQFSANFVGAGIQPLTLDGFKSQQNMWAGAPASAAAGTIDGKLPSP 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL---QKG----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSAGVAGSAGVAGSAGVAGSAGVAGSAGVAGS 1372
QY 402 YNSQVDFYNIPLARFTSEDLRRGNHNSATLDGSGPEGLPQTSNYKLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATSS---RVNITG-----DLITINGL--- 1423
QY 462 SNVDMREVMNMIIOQGFOMNKSVTATDTMLQKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVEAKR 1461

RESULT 16
US-08-302-832-2
Sequence 2, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstreser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match      6.7%; Score 173.5; DB 1; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKTHSTGLGTVSNINAN-TIGYKQOVVFDLPSQDLAIGSTGSGPNQAGCAQVGS 70
Db 1011 IKAGVDGSDSDATNNANLTKTKELK-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTITQGAPEPNS-----VTDLAIGKGFFQVLTEDKVHYTRAGNFRFTQDGFND 123
Db 1060 DLTIGNTSADGTNAKVTFNQVKDSKISADG-HKVTLSKVE--TSGSNNTEDSSDN 1116
QY 124 PSQFTLMGSRISNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSDTKTQS 183
Db 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGETITTKTGTINATTGNVEITAGTGS 1170
QY 184 EANPYFALLESKWNGCT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSG 238
Db 1171 -----ILGIESSGSVTLTATEGALAVSNIS-----GNVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVAMNPSDEGSAAGTDSAGLL--MSGTMTFSSNGELKNMTAFTPTGSATKOL 295
Db 1215 GSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLTQSNKIKATTGEANVTSATGTI 1274
QY 296 NAMQAPLVNGLPQFSANFVGAGIQPLTLDGFKSQQNMWAGAPASAAAGTIDGKLPSP 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPIQTSGNSTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL---QKG----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSAGVAGSAGVAGSAGVAGSAGVAGSAGVAGS 1372
QY 402 YNSQVDFYNIPLARFTSEDLRRGNHNSATLDGSGPEGLPQTSNYKLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATSS---RVNITG-----DLITINGL--- 1423
QY 462 SNVDMREVMNMIIOQGFOMNKSVTATDTMLQKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVEAKR 1461

RESULT 17
US-08-530-198-2
Sequence 2, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
```

CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTSTGLGTVSNNIAN-TIGYKQOVVFDLFSQDLAIGSTGSGPNQAGMGAVGS 70
DB 1011 IKAGVDGSDSDATNNALTIKTKEKL-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTFTQGAPEGNS-----VTDLAIGKGFQVTLKDVHYTRACNFRFTQDGFND 123
DB 1060 DLITGNTNSADGNAKVTNFQVQKDSKISADG-HKVTLHSHKVE--TSGSNNTEDSSDNN 1116
QY 124 PSQFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKIQS 183
DB 1117 -AGLTIDAKNVVNNITSHKAVSI-----SATSGETITKTGTINATTGNVEITAGTGS 1170
QY 184 EANPYFALLSWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTG 238
DB 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVMNPSBDGSAAGTDSAGLL---MSGTMTFSSNGELKNTAFTPTGSA TKDL 295
DB 1215 GSTIKGTESVTTSSQSGDIGTISGGTVEVKATESLTQSNKIKATGTEANVTSGTGI 1274
QY 296 NAMQAPLVNGLPOFSANFVAGIQPLTLDGFKSQONWAGAPASAAIGTDIGKLP SM 355
DB 1275 GGTISGNTVN-----VTANAGDLTVNGGAE-----INATEGAATLTSSGKL--- 1316
QY 356 MPQTSGNSTARGSSSTRYSQDGPQGLL---VDVTITSEGL---OGK----- 401
DB 1317 -----TTASSHITSAGQVNLQAGDSVAGSINAVNTLTGTLTVKGSNINATSGTL 1372
QY 402 YNSQVVDFFNYIPLARFTSDGLRREGNHYSATLDSGGPEGLPGTSGYKLSVNLQET 461
DB 1373 VINAKDAELNGAALGNHTVNNATNANGSGSVIATTSS---RVNITG---DLITINGL-- 1423
QY 462 SNVDMSEMNIIIQGFQWMSKSVTTADTMQKALELKR 502
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEIVEAKR 1461

RESULT 18

US-08-469-880-2

; Sequence 2, Application US/08469880

Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: Of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MJS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTSTGLGTVSNNIAN-TIGYKQOVVFDLFSQDLAIGSTGSGPNQAGMGAVGS 70
DB 1011 IKAGVDGSDSDATNNALTIKTKEKL-----TQDLNI-----SGFNKAEITAKDGS 1059
QY 71 VRTFTQGAPEGNS-----VTDLAIGKGFQVTLKDVHYTRACNFRFTQDGFND 123
DB 1060 DLITGNTNSADGNAKVTNFQVQKDSKISADG-HKVTLHSHKVE--TSGSNNTEDSSDNN 1116
QY 124 PSQFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKIQS 183
DB 1117 -AGLTIDAKNVVNNITSHKAVSI-----SATSGETITKTGTINATTGNVEITAGTGS 1170
QY 184 EANPYFALLSWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTG 238
DB 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214
QY 239 SKTFEYLVMNPSBDGSAAGTDSAGLL---MSGTMTFSSNGELKNTAFTPTGSA TKDL 295
DB 1215 GSTIKGTESVTTSSQSGDIGTISGGTVEVKATESLTQSNKIKATGTEANVTSGTGI 1274
QY 296 NAMQAPLVNGLPOFSANFVAGIQPLTLDGFKSQONWAGAPASAAIGTDIGKLP SM 355

Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL----- 1316
QY 356 MPIQTSSGNSSTARNGSSSTRYSODGYPOGDL--VDVTITSEGL---QKG----- 401
Db 1317 -----TTEASHITSAKQOVNLSAQDGSVAGSINAANVTLTGTLTTVKGSSINATSGTL 1372
QY 402 YNSQVDFVNIPLARFTSEGLRREGNNHYSATLDSGGPFGLPCTSNYGLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVTVNATNGSSVIATSS-----RVNITG-----DLITINGL-- 1423
QY 462 SNVDMREMVMNIIQRFQOMNSKSVTTADTLMLOKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVEAKR 1461

RESULT 19
US-08-728-470-2
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
US-08-728-470-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;
QY 12 MKTHSTGLTGVNNANAN-TIGYKQQVVFQDLFSODLAIGTSQSPNQAGMGAQVGS 70

Db 1011 IKAGVDGNSDSDATNNANLTIKTTELK-----TQDLNI-----SGPNKAEITAKDGS 1059
QY 71 VRTTFOGAPEPGNS-----VTDLAIGKGQFQVLTEDKVHYTRAGNFRFTODGFLND 123
Db 1060 DLITGNTNSADGTNAKVTFNQVKDSKISADG-HKVTLSKVE--TSGSNNTEDSDNN 1116
QY 124 PSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSGTDKTQS 183
Db 1117 -AGLITDAKNVTVANNITSHKAVSI-----SATSGETTKTGTINATTGNVEITAGTGS 1170
QY 184 EANYFALLSFKWNGT-----PPSTSNYSYAOQPMRVYDQOQNSHDITVYFDGAPSGSTG 238
Db 1171 ----ILGIESSSGSVLTATGALAVSNIS-----GNTVTVTAN--SGALATLA 1214
QY 239 SKTFEYLNAMNPSEDGSAAGTDSAGLL---MSGTMTFSSNGELKNWTATPTGTSANKDL 295
Db 1215 GSTIKGTESVTTSSQSGDIGTISGGIVFVKATESLTQSNKIKATGGEANTVSATGTI 1274
QY 296 NAWQPAFLVNLGPQFSANFVAGIQTLDLFGIKSQOMMWAGAPASAAAICTDIGKLPSM 355
Db 1275 GGTISGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL-- 1316
QY 356 MPIQTSSGNSSTARNGSSSTRYSODGYPOGDL--VDVTITSEGL---QKG----- 401
Db 1317 -----TTEASHITSAKQOVNLSAQDGSVAGSINAANVTLTGTLTTVKGSSINATSGTL 1372
QY 402 YNSQVDFVNIPLARFTSEGLRREGNNHYSATLDSGGPFGLPCTSNYGLSVNOLET 461
Db 1373 VINAKDAELNGAALGNHTVTVNATNGSSVIATSS-----RVNITG-----DLITINGL-- 1423
QY 462 SNVDMREMVMNIIQRFQOMNSKSVTTADTLMLOKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVEAKR 1461

RESULT 20
US-08-617-697-2
; Sequence 2, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810

```

; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-617-697-2

Query Match 6.7%; Score 173.5; DB 2; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKTHSTGLGTVSNNIANAN-TIGYKQQVVFQDLFSDLAIGSTGSGPNQAGMGAVGVS 70
DB 1011 IKAGVDGSDSDATNNANLTIKTKEKL-----TQDLNI-----SGFNKAETAKDGS 1059

QY 71 VRTFTQAFEPGNS-----VTDLAIGKGFQVLTEDKHYVTRAGNFRFTQDGLND 123
DB 1060 DLITGNTNSADGNTAKKVTNFQVDSKISADG-HKVTLSKVE--TSGNNNTEDSSDNN 1116

QY 124 PSFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTS 183
DB 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGEITTKTGTINATTGNVEITAGTS 1170

QY 184 BANPYFALLESWKNGT-----PPISYNSYVYQPMRVYDQGNSHDITVYFDGAPSGTG 238
DB 1275 GGTTSGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316

QY 356 MPIQTSSGNSTARNSSSTRYSQDGYPOGDL--VDVTITSEGL---OQK----- 401
DB 1317 ----TTEASHHITSKQVNLQAQDGSVAGSINAANVTNTLTGTLTVKGSINATSGTL 1372

QY 402 YNSQVVDVFNPIPLARTSEDLRREGNNHYSATLDSGGPEFGLPGTSYNGKLSVNQLET 461
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-719-641-2

Query Match 6.7%; Score 173.5; DB 3; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MKTHSTGLGTVSNNIANAN-TIGYKQQVVFQDLFSDLAIGSTGSGPNQAGMGAVGVS 70
DB 1011 IKAGVDGSDSDATNNANLTIKTKEKL-----TQDLNI-----SGFNKAETAKDGS 1059

QY 71 VRTFTQAFEPGNS-----VTDLAIGKGFQVLTEDKHYVTRAGNFRFTQDGLND 123
DB 1060 DLITGNTNSADGNTAKKVTNFQVDSKISADG-HKVTLSKVE--TSGNNNTEDSSDNN 1116

QY 124 PSFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTS 183
DB 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGEITTKTGTINATTGNVEITAGTS 1170

QY 184 BANPYFALLESWKNGT-----PPISYNSYVYQPMRVYDQGNSHDITVYFDGAPSGTG 238
DB 1171 ----TTEASHHITSKQVNLQAQDGSVAGSINAANVTNTLTGTLTVKGSINATSGTL 1214

QY 239 SKTPEYLVANMPSEDSNAAGTDSAGLL--MSGTMTFSSNGELKNTAFTPTGSATKDL 295
DB 1215 GSTIKGTESVTTSSQSGDIGTISGGIVEKATESLTQNSKIKATTGEANVTISATGI 1274

QY 296 NAWOPAPLVNGLPOFSANFVGAGIQPLTDFGKISQONMWAGAPASAAAIGTDIGKLPMS 355
DB 1275 GGTTSGNTVN-----VTANAGDLTVNGAE-----INATEGAATLTSSGKL--- 1316

QY 356 MPIQTSSGNSTARNSSSTRYSQDGYPOGDL--VDVTITSEGL---OQK----- 401
DB 1317 ----TTEASHHITSKQVNLQAQDGSVAGSINAANVTNTLTGTLTVKGSINATSGTL 1372

QY 402 YNSQVVDVFNPIPLARTSEDLRREGNNHYSATLDSGGPEFGLPGTSYNGKLSVNQLET 461
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

;
; RESULT 22
; US-09-206-942-67
; Sequence 67, Application US/09206942

```

Patent No. 6432669
GENERAL INFORMATION:
APPLICANT: Loomsore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
FILE REFERENCE: 1038-861 MIS-jb
CURRENT APPLICATION NUMBER: US/09/206,942
CURRENT FILING DATE: 1998-12-08
EARLIER APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 67
LENGTH: 1536
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-206-942-67

Query Match 6.7%; Score 173.5; DB 4; Length 1536;
Best Local Similarity 22.8%; Pred. No. 3.6e-06;
Matches 119; Conservative 80; Mismatches 222; Indels 100; Gaps 24;

QY 12 MTHSTGLGTNNANAN-TTGYKQQVVFQDLFSQDLAIGSTGSGPNOAGMGAQVGS 70
DB 1011 IRAGVDGNSDGNANLTIKTKELK-----TQDLNI-----SGFNKAEITAKDGS 1059

QY 71 VRTITQGAFFGNS-----VTDLAIGGKFFQVLTEDKVHYTRAGNERFTQDFLND 123
DB 1060 DUTIGNTNGADGTNAKVTNFQKQKISADG-HKVTLLHSKVE--TSGNNNTDSSDNN 1116

QY 124 PSGFTLMGSKRIINNPNKIKETLEPIQLDFNDFTVAKSPAKTSTALNAVNVNLDGSDTKTQS 183
DB 1117 -AGLTIDAKXNVVNNITSHKAVSI-----SATSGEITTKGTGTINATNGVVEIQAOTGS 1170

QY 184 EAPFPALLESWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSSGTG 238
DB 1171 ----ILGGIESSGVSLLPATGALAVSNIS-----GNTVTVTAN-SGALTTLA 1214

QY 239 SKTFEYLVAMNPFSEDSAGSDSAGLL---MSGTMTFSSNGELKNMTAFTPTGSATKDL 295
DB 1215 GSTIKTESVTTSSQSGDIGITGGTVEKATESITQTSNKKIKATTGEANVTSATGTI 1274

QY 296 NAWQAPLVNGLPQSFANFVAGIQPLTLDFGKQSQNNWAGAPASAAAIGTDIGKLPSPM 355
DB 1275 GGTISGNTVN-----VTNAGDLTVNGAE-----INATEGAATLTTSSGKL--- 1316

QY 356 MPIQTSSGNSSTARNGSSSTRYSQDGYPOGDI--VDVTITSEGL---QGGK----- 401
DB 1317 ----TTEASHTSIAKQVNLGAQDSVAGSINAANVILNTTGTITVYKGSINATSGTL 1372

QY 402 YNSQVVDVFNIPFIARFTSDEGLRREGNNHYSATLDSGGPFGLPQTSNYGKLSVNQLET 461
DB 1373 VINAKDAELNGAALGNHTVNVNATNGSGSVIATTS---RVNITG---DLITINGL-- 1423

QY 462 SNVDMREVMNLIIGFQGMKSQSVTTADTLMQKALELKR 502
DB 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEVIEAKR 1461

RESULT 23
US-08-436-748-6
Sequence 6, Application US/08436748
Patent No. 5827654
GENERAL INFORMATION:
APPLICANT: CHAN, YOON LOONG
APPLICANT: LOUIE, HELENA
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
CAMPYLOBACTER
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney

STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/436,748
APPLICATION NUMBER: US/08/436,748
FILING DATE: 05-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-428 MIS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-748-6

Query Match 6.6%; Score 169.5; DB 2; Length 248;
Best Local Similarity 27.0%; Pred. No. 4.7e-07;
Matches 70; Conservative 35; Mismatches 121; Indels 33; Gaps 9;

QY 1 MMGSLFGATGMKTHSTGLTGVNANNIANTIGYKQQVVFQDLFSQDLAIGSTGSGPN 60
DB 1 MDNALYVGLSRQMTVRRELDIVANNIANTTGFKVEDLMVRT--EQAKPAKTLDGSSPV 58

QY 61 QAGMGAQVGSVRTITFTQGAFFGNSVTDLAIGGKFFQVLTEDKVHYTRAGNERFTQDGF 120
DB 59 KFMVMT---GVRRNFTQPMTKGGDYDLAINGNGFKVQANGERYTRDGRFTTNEGI 115

QY 121 LNDPSGFTLM---GSRISNNPNKIKETLEPIQLDFNDFTVAKS-----PAKST 166
DB 116 LVTQAGAPVLDGGGQITIDPR-----LGPVTVG-KDGIVSQGAIRVSRIGLVRPDDLST 169

QY 167 ALNAVNLGSDTKTQ-----SEANPYFALLESWKNGTPEISTSNYSYAQPMRVYDQGN 222
DB 170 FAKDGNLYRNTTNTAPQPVTDQAIHQGMLEA--SNVQPVIEITKLEIQ--RAYESVAK 225

QY 223 SHDITVYFDGAPSSGTGSKT 241
DB 226 MMDNTAELSRTPSSVWARS 244

RESULT 24
US-08-483-857-6
Sequence 6, Application US/08483857
Patent No. 6020125
GENERAL INFORMATION:
APPLICANT: Chan, Yoon Loong
APPLICANT: Louie, Helena
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
CAMPYLOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,857
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-857-6

Query Match 6.4%; Score 165; DB 3; Length 247;
Best Local Similarity 27.0%; Pred. No. 1.2e-06;
Matches 70; Conservative 36; Mismatches 119; Indels 34; Gaps 10;

QY 1 MMGLFICATGKMTSTGLGTVSNINANTIGYKQVQVFDLFSQDLAIGSTGSGGN 60
DB 1 MDNALYVGLSRQMTVRRELDIVANNANTIGKVEDLMVRT--EQAKPAKTLDGSSPV 58
QY 61 QAGMGAQGVSVRTFTQGAPEPGNSVTDLAIGGKGFQVTLDEKHYTRAGNFRFTQDGF 120
DB 59 KFMVMT---GVRNFTQGTGKTDGYDLAINGWGFQANGERYTRDGRFTTNPEGI 115
QY 121 LNDPSGFTLM---GSRISNNPNKKTLEPIQLDNDPTVAKS-----PAKIST 166
DB 116 LVTQAGAPVLDGGGQITIDPR-----LGEVTVG-KDGIVSQAIRVSLGLVRPDLST 169
QY 167 ALNAVNLGSDTKTQ----SEANPYFALLESKNGTPTPSTSNYSYAQPMRVVDQGN 222
DB 170 FAXDGNLXNTTAPQPVTDQAIIHQGMLEA--SNVQPVIEITKLEIQ--RAVESVAK 225
QY 223 SHLITVFDGAPSTGSKT 241
DB 226 MND-TAELSRTPSSVWARS 243

RESULT 25
US-08-436-748-3
Sequence 3, Application US/08436748
Patent No. 5827654
GENERAL INFORMATION:
APPLICANT: CHAN, VOO LOONG
APPLICANT: LOUIE, HELENA
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
CAMPYLOBACTER
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,748
FILING DATE: 05-AUG-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-428 MIS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-436-748-3

Query Match 6.0%; Score 155; DB 2; Length 270;
Best Local Similarity 30.7%; Pred. No. 1.1e-05;
Matches 46; Conservative 28; Mismatches 64; Indels 12; Gaps 5;

QY 1 MMGLFICATGKMTSTGLGTVSNINANTIGYKQVQV---FQDLFSO---DLAIGS- 53
DB 1 MONGYQATGGMVTFQFKLDVITNNLANINTSGYKRDVVIADFKRIFKETQDELPIENH 60
QY 54 --TSGQPNQAGM--AQGVSVRTFTQGAPEPGNSVTDLAIGKGFQV--TLEDKVHYT 108
DB 61 TRDASRFVNTTIDGIPVQSQBYTDFSLGSLKATNPLDLAMTREDAFYLVQTKDGEVRLT 120
QY 109 RAGNFRFTQDGLNDPSGFTLMGSRISNPN 138
DB 121 KQGNFQDLDEGLVNGKQYKVLSSDYFNNP 150

RESULT 26
US-08-483-857-4
Sequence 4, Application US/08483857
Patent No. 6020125
GENERAL INFORMATION:
APPLICANT: Chan, Voon Loong
APPLICANT: Louie, Helena
TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
CAMPYLOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,857
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-857-4

```

Query Match      6.0%; Score 155; DB 3; Length 270;
Best Local Similarity 30.7%; Pred. No. 1.1e-05;
Matches 46; Conservative 28; Mismatches 64; Indels 12; Gaps 5;

QY 1 MMGSLFIGATGKMTSTGLTGNINIANANTYKQOQVW---FQDLFSQ---DLAIGS- 53
DB 1 MONGYQATGGWVTOFNKLDVITNNLANINTSGYKRDVVIADFKRIKETOQDELPENH 60
QY 54 --TGSGPNOAGWG-AQVGSVITITQGAFFGNSVTDLAIGKGFQV--TLEKQVYT 108
DB 61 TRDASRFVNTTIDGIPQVSGEYTDPSLGLKATNPLDLAMTREDAFYLVQTKDGEVRLT 120
QY 109 RAGNFRFTODGFLNDPSGFTLMGSRISNNP 138
DB 121 KQNFOLDDEGLVNVKQGYKVLSSDYFNPP 150

RESULT 27
US-09-206-942-65
; Sequence 65, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-65

```

```

Query Match      5.8%; Score 149.5; DB 4; Length 1180;
Best Local Similarity 20.2%; Pred. No. 0.00034;
Matches 113; Conservative 88; Mismatches 212; Indels 147; Gaps 22;

QY 30 NTIGYKQQVVFQDLFSQDLAIGSGQGNQAGMGAGVGSVRTITFTQGAFFGNSVTDL 89
DB 606 NISGFKAKIVAKD--SSNLTIGNSDDSGNT-----SAKTV-----TFNNVKDS 647
QY 90 AIGCKGFQVLTEDKHYTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQ 149
DB 648 KISADG-HKVTLSKV-K-TLSDNNDNTEGSDNN-TGLTITAKDVEVNNIT----- 696
QY 150 LDFNDPTVAKSPA-----KTSTALNAVNLGSDTDKTSQ-----EANFYALLESW--- 195
DB 697 ---SHKTVNVSANGGITTKTGTINATAGNVEITAGTSIOGGIESKPGSVTIAGGDT 753
QY 196 -----KGN-----GTPPISTSNYSYAQPMRVYDQGN----- 222
DB 754 LAVGNISGNAVTVTANGSALTTLAGSTIKGTESITSS-----QSGNIGGKISG 802
QY 223 -----SHDITVYFDGAPSGTKTFEYLVAAMPSEDGSAAG-----TDSAG 264
DB 803 KTVNVTATSLTTQADSKIEATG---EANTVSKTSIIIGTISGTVTEATEGLTQAG 859
QY 265 LMSGTMTFSSNGELKNMTAFTPTG-----SATKDL-----NAWQAPLVN 305
DB 860 STITGTESVTSSQSGNIGMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTD 919
QY 306 GLPQSFANFVAGIQLPTLDGFKISQONMWAGAPASAAAGTIDGKLPMSWPIQTSSGNS 365
DB 920 GTISGNTVNTANGDITVDEDAK-----IDATGGAATLTATSGKL-----TTKASS 966
QY 366 TARGSSSTRYSQDGPQDGL--VDVTITSEGLQKYSNSQVDFYNIPLARFTSEGD 423

```

```

DB 967 SITSAANQVNLKADSGISGINNAANVTNTGALTIVKSSINANSGLVINAKDAELN 1026
QY 424 LRREGNNHYSATLDSGGPEGLPGTSGNYKLSVNOLETNSVD-MSREWNMIILQRFQFM 482
DB 1027 GEASGHTVNVATNANGSGSVIATTSRVNITGDLITINGLIISKNGINTVILL-KGVKI 1085
QY 483 NSKSVTTADTMLKALELKR 502
DB 1086 DVKIQPGIASVDEVEAKR 1105

RESULT 28
US-09-206-942-63
; Sequence 63, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-63

```

```

Query Match      5.8%; Score 149.5; DB 4; Length 1188;
Best Local Similarity 20.2%; Pred. No. 0.00034;
Matches 113; Conservative 88; Mismatches 212; Indels 147; Gaps 22;

QY 30 NTIGYKQQVVFQDLFSQDLAIGSGQGNQAGMGAGVGSVRTITFTQGAFFGNSVTDL 89
DB 614 NISGFKAKIVAKD--SSNLTIGNSDDSGNT-----SAKTV-----TFNNVKDS 655
QY 90 AIGCKGFQVLTEDKHYTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQ 149
DB 656 KISADG-HKVTLSKV-K-TLSDNNDNTEGSDNN-TGLTITAKDVEVNNIT----- 704
QY 150 LDFNDPTVAKSPA-----KTSTALNAVNLGSDTDKTSQ-----EANFYALLESW--- 195
DB 705 ---SHKTVNVSANGGITTKTGTINATAGNVEITAGTSIOGGIESKPGSVTIAGGDT 761
QY 196 -----KGN-----GTPPISTSNYSYAQPMRVYDQGN----- 222
DB 762 LAVGNISGNAVTVTANGSALTTLAGSTIKGTESITSS-----QSGNIGGKISG 810
QY 223 -----SHDITVYFDGAPSGTKTFEYLVAAMPSEDGSAAG-----TDSAG 264
DB 811 KTVNVTATSLTTQADSKIEATG---EANTVSKTSIIIGTISGTVTEATEGLTQAG 867
QY 265 LMSGTMTFSSNGELKNMTAFTPTG-----SATKDL-----NAWQAPLVN 305
DB 868 STITGTESVTSSQSGNIGMISGGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTD 927
QY 306 GLPQSFANFVAGIQLPTLDGFKISQONMWAGAPASAAAGTIDGKLPMSWPIQTSSGNS 365
DB 928 GTISGNTVNTANGDITVDEDAK-----IDATGGAATLTATSGKL-----TTKASS 974
QY 366 TARGSSSTRYSQDGPQDGL--VDVTITSEGLQKYSNSQVDFYNIPLARFTSEGD 423
DB 975 SITSAANQVNLKADSGISGINNAANVTNTGALTIVKSSINANSGLVINAKDAELN 1034
QY 424 LRREGNNHYSATLDSGGPEGLPGTSGNYKLSVNOLETNSVD-MSREWNMIILQRFQFM 482

```

Db 1035 GEASGHTVNVATNANGSGSVIATTSRNVITGDLITINGLIISKNGINTVLL-KGVKI 1093

Qy 483 NKSQVTTADTLMQALELKR 502
Db 1094 DVKVIQPIASVDEIAKR 1113

RESULT 29

US-09-538-092-330
; Sequence 330, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 330
; LENGTH: 1306
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YGR014W
US-09-538-092-330

Query Match 5.5%; Score 143; DB 4; Length 1306;
Best Local Similarity 19.6%; Pred. No. 0.0015;
Matches 122; Conservative 75; Mismatches 245; Indels 182; Gaps 25;
Qy 4 SLFICATGMKTHSTGLGTGVNNIANANTIGYKQOVVFO--DLFSODLAIGS-TGSQGN 60
Db 598 SPYTAGASTEASL--ISSTAETSQVSOSTTALQTSFASSTTEGSETSSQGPS 655
Qy 61 QAGGAQVGSVRIFTGAGEPGNSVTDL-----AIGKGFFQVLEKDVHYTRAG 111
Db 656 TSSVLVQMPs-----STSEFSPQTTQMNASASSSQYTISSTGILSQVSDTSVSYTSS 711
Qy 112 NFRFTQDGLNDPSGFTLMGSRIS-----NNPNIKETLEPIQLDFNDPTVAK- 159
Db 712 S--SVSQVSDTPVSYTTSSSVSQVSDTPVSYTTSSSVSQVSDTPVSYTTSSSVSQV 768
Qy 160 --SPAKTSTALNAVNLGSDTDTQSEANPYFALLESWKNGTTPPISTSNYSVA----- 211
Db 769 SDTPVSYTTSSSVSQVSDTSVPSTSSRSVQV-----SDTPVPSSTSSRSVQVSDTS 822
Qy 212 -QMRVVYDQ-----GRSHDI-----TVYED 231
Db 823 LQPTTSSQFTTSTHCALESSESVSQASEITSSINATASEVHSIQTTAATQSTLTST 882
Qy 232 GAPSSSTGSKTFEYLVAMNPESDGAAGTDSAGLMSGTMTFSSNGELKNMTAFTPTGSA 291
Db 883 DANSSASAPLE--VATSTPTPSKASSL-----LLTPSTSSLSQVATNNVQTSLTTEST 936
Qy 292 T-----KLANAWQAPLVNGLPQ--PSANFVGAGIQLTL----- 324
Db 937 TVLEPSTTSSSTFSLVSSDNNWIPTELITQAPAASTASSTVGGTQMTLPLRAIAA 996
Qy 325 -----DFGIKSOON--MWAGAPASAAIGTDIGKLP----- 353
Db 997 TQVPEPEGVTLITIGFKALNYEFVSEPKSAQI--FCYLPEALNTFPKVNFTNITVL 1053
Qy 354 SMPIQTSSGN-----STAENGSSSTRRYSQDY-----PQGLDVT 391
Db 1054 QIVPLQDDSLNVLVSVAEVYFPTAEIEELSNLTNSSA--FYTDGWTAKMAAWDSS 1111

Qy 392 ITSEBKLOGKYNSQVDFYNIPLARFTSEGLRRRGNHYSATLDSGGPEFGPLGTSNY 451
Db 1112 IPLTGLLHDSNSNS--GGSDGSSNSNSGSGSGNSNSGVSSSGNSYQDAGTLEY 1168
Qy 452 GKLSVNOLETNSVDMREMANII 475
Db 1169 SKKSNVSTSS-KSKKIIIGLVI 1191

RESULT 30

US-09-206-942-45
; Sequence 45, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS,jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-45

Query Match 5.5%; Score 142; DB 4; Length 1095;

Best Local Similarity 20.9%; Pred. No. 0.0014;
Matches 101; Conservative 76; Mismatches 229; Indels 78; Gaps 18;

Qy 30 NTIGYKQOVVFDLFSODLAIGSTGSGPNQAGMGAQVGSVRTIFTQGAFFPGNSVTDL 89
Db 604 NISGFKRAEITAKE--GADLIIGN--SDNNNA-----NAKKVTF-----NOVKDS 645
Qy 90 AIGKGFFQVLEKDVHYTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKETLEPIQ 149
Db 646 KISADS-HNVTLSKVTSSNGNDABSNNG---DGTSLTINAKNITVNNITSHKVTNIT 701
Qy 150 LDFNDPTVAKSPAKTSTALNAVNLGSDTDTQSEANPYFALLESWKNGT-----PPIS 204
Db 702 ASENVT-----KAGTINATTGSEVETAKT---GDIKGVESTSGSVTLTATGEALA 751
Qy 205 TSNSYSAQPMRVYDQGNSHDIIVYFDGAPSSSTGSKTFEYLVAMNPESDGAAGTDSAG 264
Db 752 VSNIS-----GNTVTITANKGLTQAGS-TVSAINGVTASSQSGDISGITSN 799
Qy 265 LL--MSGTMTFSSNGELKNMTAFTPTGSAKOLINAWQAPLVNGLPQFSANFVGAGIOP 321
Db 800 TVKYSAGDLTTKSGSEIKAKTGEANTVSATGTIGTISGNV-----VTANTGD 850
Qy 322 LTLDFGKSOONWAGAPASAAIGTDIGKLPSPMPTQSSNGSTARNSSSTRYSQDG 381
Db 851 LTVEDAAK-----IDATGAATLTATSGKL-----TTKASSITSNANNVLSAKDG 897
Qy 382 YPOGDL--VDVTITSEKLOGKYNSQVDFYNIPLARFTSEGLRRRGNHYSATLDSG 439
Db 898 SIGNINAAVNTLTTCALTIVKSSINANSGLVINAKDAELNGEASGHTVNVNATNAN 957
Qy 440 GPPEGLPGTNGYKLSVNOLETNSVD--MSREMNIIIQRFQMNKSVSTTADTLMQAL 498
Db 958 GSGSVIATTSRNVITGDLITINGLIISKNGINTVLL-KGVKIDVKYIQFIQASVDEVI 1016
Qy 499 ELKR 502
Db 1017 EAKR 1020

RESULT 31
US-09-206-942-43
; Sequence 43, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MLS:Jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1101
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-43

Query Match 5.5%; Score 142; DB 4; Length 1101;
Best Local Similarity 20.9%; Pred. No. 0.0014;
Matches 101; Conservative 76; Mismatches 229; Indels 78; Gaps 18;

QY 30 NTIGYKQQVVFODLFSQDLAIGSTGQGNPQAGMGAQVGSVRTIPTQGAFFGNSVTDL 89
Db 610 NISGPKAEITAKE--GADLIIGN--SDNNNA-----NAKKVTF-----NQVKDS 651
QY 90 AIGGKFFQVTLBDDKHYHTRAGNFRITQDGLNDPSGFTLMGSRISNNPNIKETLEPTQ 149
Db 652 KISADS-HNVTLSKSVETSGNNDAESNNG---DGTSLTINAKNHTVNNNITSHKTVNIT 707
QY 150 LDENDPVAKSPAKTSTALNAVNLGDSTDKTQSEANPYFALLESWKNGT-----PPIS 204
Db 708 ASENVT-----KAGTTINATGSEVTAKT-----GDKGVESGSGVTLTARGEAL 757
QY 205 TSNSYQAQPMRVYDQGNSHDITVYFDGAPSGTGTFFYLVAAMPSEDEGSAAGTDSAG 264
Db 758 VSNIS-----GNTVTITANKGLTTQAGS-TVSAINGVTATSSQSGDISGTISGN 805
QY 265 LL---MSGTMTFSSNGELKNMTAFTTGSATKDLNAWQAPLVNGLPQFSANFVGAGIOP 321
Db 806 TVKVAIGDLTTKSGSEIKAKTGEANVTSAITGTTISGNV-----VTANTGD 856
QY 322 LTLDGFIKSOQNMWAGAPASAAAIGTDIGKLPMMPIQTSSGNSTARNSSSTRYSODG 381
Db 857 LTVEDAAK-----IDATGGAATLTATSGKL-----TTKASSITSANNQVNLAKDG 903
QY 382 YPOGDL--VDVTTSEGLKQKYSNQVDFYNIPLARFTSEGLREGNNHYSATLDSG 439
Db 904 SIGGNINAAAVTLTTTGALTTVKGSINANSGLTVINAKDAELNGEASGHTVTVNATNAN 963
QY 440 GPERGLPTSNYKLSVNOLETNSVD--MSREVMNIIIOGFQMGKSVTATDTMLQKAL 498
Db 964 GSGSVIATTSRNVITGDLATINGLNIISKNGINTVLL-KGVKIDVKYIOPGIASVDEV 1022
QY 499 ELKR 502
Db 1023 EAKR 1026

RESULT 32
US-09-381-656-1
; Sequence 1, Application US/09381656
; Patent No. 6645509
; GENERAL INFORMATION:
; APPLICANT: SERRE, Guy Bruno Rene
; APPLICANT: SIMON, Michel
; APPLICANT: WEBER-VIVAT, Marina

; TITLE OF INVENTION: POLYPEPTIDE EXPRESSED IN THE HORN Y LAYER OF EPIDERMIS
; TITLE OF INVENTION: AND USE THEREOF
; FILE REFERENCE: 016800-336
; CURRENT APPLICATION NUMBER: US/09/381,656
; CURRENT FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: FR 97/03899
; PRIOR FILING DATE: 1997-03-28
; PRIOR APPLICATION NUMBER: FR 97/11317
; PRIOR FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Human
US-09-381-656-1

Query Match 5.4%; Score 140.5; DB 4; Length 529;
Best Local Similarity 20.3%; Pred. No. 0.0062;
Matches 108; Conservative 60; Mismatches 182; Indels 183; Gaps 24;

QY 10 TGMKTHSTGLTGVSNNTIANANTIGYKQOQVVFQDLFSQDLAIGSTGQGNPQAGMGAQV 69
Db 57 TG-KGDSGSGFSSYSGSSSGSSSISSARS-----SGGSGSSSGSSSIAGQG 101
QY 70 SVRTIFTQGAFFGNSVTDLAI-----GKGFQV-----TLEDKVHY 107
Db 102 SA-----GSPKPTGTGVSQSYSGSSGSLQAGSSQLGSSSHSGSSSHSGSSSHS 155
QY 108 TRAGNFRITQDGLNDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTA 167
Db 156 SSSSSQFQSSSSFO-----VGNLSALPTNDNSYRGILNPSQ-----PGOSSSSSOTS-- 202
QY 168 LNAVNLGDSIDTKQ-----SEANPYFALLESWKNGTTPPISTSNYS 209
Db 203 --GVSSSQSVSSNQRPCSDIPSPCSGGPIVSHSGPYIPSSHVSGGQRPVV----- 255
QY 210 YQPMRVYDQGNSHDITVYFDGAPSGTGT--SKTFFYLVAAMPSEDEGSAAGTDSAGLL 266
Db 256 -----VDQHGSGAPGV--QGPPCSNGGLPGKPCPPITSDVKSYGYEVVGGSSDYL 307
QY 267 MSGTMTFSSNGELKNMTAFTTGSATKDLNAWQAPLVNGLPQF----- 310
Db 308 VPG-MTYSKG-----KIYVGYFTKE-NPVKGS---GVPSAAGPPISEGYFSSNP 355
QY 311 -----SANFVGAGIOPLTDFGIKSOQNMWAGAPASAAAIGTDIGKLPMMPIQTSSGN 365
Db 356 IIPSSAASSAIAAPV-----GTGVQLCGGSGTGKPCSPSSSRVPSSSSISSAGSP 411
QY 366 TARNSSSTRYSODGYPOGDLVDVTTSEGLKQKYSNQVDFYNIPLARFTSE--- 422
Db 412 YHPCGSAQSQSPSPG-----TGSFSSSS-----SSQSSGK 442
QY 423 -----GLRREGNNHYSATLDS-----GGPEFGLP-----GTSNYGKL 454
Db 443 ILQPCGSKSSSGHPCMSVSLTLTGPPD-GSPHPDPSAKAPCGSSSAGKI 494

```
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Xenopus laevis
US-09-479A-20

Query Match          5.3%; Score 137; DB 4; Length 585;
Best Local Similarity 20.8%; Pred. No. 0.0015;
Matches 108; Conservative 62; Mismatches 191; Indels 158; Gaps 23;

QY 10 TGMKTHSTGLTGVNNIANANTIGYKQOVVFDLFSQDLAIGSTGSGGNQAGMGAQVG 69
Db 122 TGETNRATAPGTAANGSPAA-----PDLISGFGLSGLGNLGMG----- 160
QY 70 SVRIIFTQGAPEPGNSVTDLAIGCKGFQVLTEDKVVHYTRAGNFRFTQDGFNDPSGFTL 129
Db 161 -----SNFWELOOQORQUMSNPEMLSQIMENPLVQNMMSNPDLMRQMIIANPQMQL 214
QY 130 MG-----SRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTAL-----NAVNLGD----- 176
Db 215 MERNPEISHMLNPELAKQWE-----LARNPAMQEMMRNQDRALSULESIPGG 264
QY 177 -----STDKTQSEANPYFALLESWKNGTTPPISTN-----YSYAQP 213
Db 265 YNALRRMYTDIOEPMFSAAREQFGNPFSAAGSGSASOPLRTENRPLPNPWPSPASP 324
QY 214 MRVYDQOQNSHDITVYFDGAPSSGSKTFEYL-VAMNPSDGGAAAGT-----DSAGLL 266
Db 325 SS-QNQTNSB-----SNTGSTTSQSVPTVSNPLGNAASLGTGYNSPENGQLL 373
QY 267 MSGTNTSSNGELKNMFAFTPTGSAKDLNAMOP-----APLVNGLPQFSAN----- 313
Db 374 QQ-----ITENPQIQSMISAPYTRSMQMAONPEFTAQMGNIPIFSGNPQLQELRHQ 429
QY 314 --FVGAGIQP-----LTDFGIKSOQNMWAGAPAGAAAGTIDIGKLP 355
Db 430 LPVFLQMQNPESVMNSNPRAMQALLOVQOGLTQTEAPGLLSLGVG-----IPGV 484
QY 356 MPIOTSGNSTARNSSSTARY-----SODGYPOGLDVDTTITSEGLQKYSNSQV 408
Db 485 PP--TSGGSTAPENPASSSTPSSASPSSGSSNNPQQMMQMIQ---LLAG--GNSQVQ 537
QY 409 DFYNIPLARTSE-----DGLRREGNNHYSATLDSGG 440
Db 538 N-----PEVRFQSDQLDNAMGFNREAN--VQALITGG 570

RESULT 34
US-09-492-709A-302
; Sequence 302, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Orlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 302
; LENGTH: 2383
; TYPE: PRT
; ORGANISM: E. Coli

Query Match          5.3%; Score 136.5; DB 4; Length 2383;
Best Local Similarity 20.9%; Pred. No. 0.015;
Matches 113; Conservative 79; Mismatches 217; Indels 131; Gaps 25;

QY 16 STGLGTGVNNIANANTIGYKQOVVFDLFSQDLAIGSTGSGGNQAGMGAQVGSVRT-- 73
Db 1505 ATVADEGNAIANT-----EVTF--TLPEDEVKANFTLSG-----GKVIIDA 1544
QY 74 -----IPTQGAPEPGNSVTDLAIGCK-----GPFQVLTEDKVVHYTRAGNFRFTQDGFNL 122
Db 1545 EGKAKVTLKGTGAGAHVTASMTGKSEQLVNVFIADTLTAQV-----NLNVTFEDNFIA 1598
QY 123 DPSGFTLMSGISIN-NNPIKKEKLEPIQLDFNDPTVAKSPAKTSTAL-----NAVNLG 175
Db 1599 NNVGNMTRLCATVTDGNGN-----PLA--NEAVTFLPADVSASFLLGGGSAITDIN 1648
QY 176 DSTDKTQSEANPYFALLESWKNGTTPPISTNYSYAQPMRVYDQOQNSHDITVYFDGAPS 235
Db 1649 GKAEVTLSGTK-----SGTYPVTVSNNYG-----VSDTKQVTLIADAGTA 1689
QY 236 STGSKTFEYLAMNPSDEGS--AASGDSAGLLMSG-----TMTFSS-----NGEL 279
Db 1690 KLASLTSYVSFVSTTEGATMTASVTDANGNPVEGKVNFRGTSVTLSSTSVETDDRGPA 1749
QY 280 KNMTAFTPTGSAKDLN-AWQAPLVNGLPQFSANFVAGIQLTL-----DFGIK 329
Db 1750 ELVLTSTEVGLKTVSASLADKPTFVISRLNASADVNSATITSLIPEGVVMAQDVAVK 1809
QY 330 SQQNMWAGAPASAAAGTIDICKLPSPMPLOTSSGNS-----ARNGSSSTRYSQ 379
Db 1810 AHVNDQFNPVHQPVTFSAPSSQMIISQNTVGTQGVAEVMTPTERNGSYMVKASLP 1869
QY 380 DG--YPOGLDLD--VTITSEGLQKYSNSQVDFYNIPLARTSEDLRREGN-NHYS 433
Db 1870 NGASLEKLEALDEKLTJTASSPLIGVYAPTGA-----LTATLTSANGTPVEQVINFS 1924
QY 434 ATLSGGEPEGLPGTSYNGKLSVNOLETNSVDMREVMNMIIQRFQMSKSVTTADTM 493
Db 1925 VT-----PE--GATLGGKVRTNSGQAPVVLTSNKVGTVYTTASPH--NGVTIQTQTTV 1975

RESULT 35
US-08-621-944A-4
; Sequence 4, Application US/08621944A
; Patent No. 6440425
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER
; MEMBRANE PROTEIN OF MORAXELLA
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sim & McBurney
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621,944A
; FILING DATE: 26-MAR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,370
```


FILING DATE: 07-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24, 973
REFERENCE/DOCKET NUMBER: 1038-587
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 593-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1833 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-621-944A-4

Query Match 5.1%; Score 133; DB 4; Length 1833;
Best Local Similarity 20.0%; Pred. No. 0.021;
Matches 126; Conservative 79; Mismatches 222; Indels 204; Gaps 29;

```
QY 7 IGATGKTHSTGLTGVNNANANTIGYKQOVVFQDLFSQDLAIGSTGSGQ-----PN 60
DB 506 VGANGIKFTNVNGSNPTGIANTARI-----TRD-KIGFAGSDGAVDTNKPY 551
QY 61 QAGMGAQGVSVRTITQGAPEPGNSVTDL-----AIGGKGFQV-----TLEDKVHYTRAG 111
DB 552 LDQDKLQGVNVK-ITNTGINAGGKAITGLSPTLPSIADQSSRIELGNTIQDKKSNAA 610
QY 112 NFRTOQDFLND--PSGFTLMGSRISNPN-----138
DB 611 -----INDILNTGFNL-----KNNNPIDFVSTVDIVDFANGNATTATVTHDTANKTS 658
QY 139 -----NIKETLEPIQLDFN-----DPTVAKSPAKTSTALN-----AVNLDGS 177
DB 659 KVVYDVNVDDTTHLTGTDNKKLGKVTTKLNKTSANGNTATNFVNSDEDAVNAKDI 718
QY 178 TDKTQSEANPYFALLESWKNGTPPISTSNYSVAQPMRVYDQGNSHD---ITVYFDGAP 234
DB 719 AENLNTAKE-----IHTTKGTADTALQFT-----VKKVDENNADANAITVQKQAN 768
QY 235 SSTGSKTFEYLAVNPFSD--GSAASGTDGAGLLMSGTMTPFSNG--ELKMTAFTPTGSA 291
DB 769 NQVNTLTLKGNGLNITKDXGTVTFGINTTSGLKAGKSTLNDGGLSIKN-----PTGSE 823
QY 292 TKDLNAPQAPLVNGLPQFSAN---FVGAGIQPLT-----323
DB 824 QIQVGA-----DGKFAKVNNGVVGAGIDGTRITRDEIGFTGTGSLDKSPHLK 876
QY 324 --LDFGIKSOQNMWAGAPASAAAIGTDIGKLPSPMPITSSGNSTARNSSSTRYSQDG 381
DB 877 DGINAGGKKTNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSAKTAKNSLHEFSV-A 935
QY 382 YPQGLVDVTTITSEGLQGYKSNQVDFYNIPLAFRTSBDGLRRE-----427
DB 936 DEQGN--NFTVSNP---YSSYDTSKTSN-----VITPAGENGITTKVNGVVRVGDQTK 985
QY 428 -----GNHYSATLDSGGPEFGLPQTSN-----YKLSVNOLETSNV---DMS 467
DB 986 GLTTPKLTGVNNGKGVIVDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDEK 1045
QY 468 REMVNMIIQGFQMNKS-----VTTADTM 493
DB 1046 RAASIVDLVLSAGFNLOQNGEAVDFVSTYDTV 1076
```

RESULT 36
US-08-945-567D-4
Sequence 4, Application US/08945567D
Patent No. 6448386
GENERAL INFORMATION:
APPLICANT: SASAKI, Ken
APPLICANT: HARKNESS, Robin E.

APPLICANT: LOOMORE, Sheena M.
APPLICANT: CHONG, Pele
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF
TITLE OF INVENTION: MORAXELLA
FILE REFERENCE: 1038-745 MIS
CURRENT APPLICATION NUMBER: US/08/945,567D
PRIOR FILING DATE: 1996-04-29
PRIOR APPLICATION NUMBER: 08/431,718
PRIOR FILING DATE: 1995-05-01
PRIOR APPLICATION NUMBER: 08/478,370
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: 08/621,944
PRIOR FILING DATE: 1996-03-26
PRIOR APPLICATION NUMBER: PCT/CA96/00264
PRIOR FILING DATE: 1996-04-29
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 1833
TYPE: PRT
ORGANISM: Moraxella catarrhalis
US-08-945-567D-4

Query Match 5.1%; Score 133; DB 4; Length 1833;
Best Local Similarity 20.0%; Pred. No. 0.021;
Matches 126; Conservative 79; Mismatches 222; Indels 204; Gaps 29;

```
QY 7 IGATGKTHSTGLTGVNNANANTIGYKQOVVFQDLFSQDLAIGSTGSGQ-----PN 60
DB 506 VGANGIKFTNVNGSNPTGIANTARI-----TRD-KIGFAGSDGAVDTNKPY 551
QY 61 QAGMGAQGVSVRTITQGAPEPGNSVTDL-----AIGGKGFQV-----TLEDKVHYTRAG 111
DB 552 LDQDKLQGVNVK-ITNTGINAGGKAITGLSPTLPSIADQSSRIELGNTIQDKKSNAA 610
QY 112 NFRTOQDFLND--PSGFTLMGSRISNPN-----138
DB 611 -----INDILNTGFNL-----KNNNPIDFVSTVDIVDFANGNATTATVTHDTANKTS 658
QY 139 -----NIKETLEPIQLDFN-----DPTVAKSPAKTSTALN-----AVNLDGS 177
DB 659 KVVYDVNVDDTTHLTGTDNKKLGKVTTKLNKTSANGNTATNFVNSDEDAVNAKDI 718
QY 178 TDKTQSEANPYFALLESWKNGTPPISTSNYSVAQPMRVYDQGNSHD---ITVYFDGAP 234
DB 719 AENLNTAKE-----IHTTKGTADTALQFT-----VKKVDENNADANAITVQKQAN 768
QY 235 SSTGSKTFEYLAVNPFSD--GSAASGTDGAGLLMSGTMTPFSNG--ELKMTAFTPTGSA 291
DB 769 NQVNTLTLKGNGLNITKDXGTVTFGINTTSGLKAGKSTLNDGGLSIKN-----PTGSE 823
QY 292 TKDLNAPQAPLVNGLPQFSAN---FVGAGIQPLT-----323
DB 824 QIQVGA-----DGKFAKVNNGVVGAGIDGTRITRDEIGFTGTGSLDKSPHLK 876
QY 324 --LDFGIKSOQNMWAGAPASAAAIGTDIGKLPSPMPITSSGNSTARNSSSTRYSQDG 381
DB 877 DGINAGGKKTNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSAKTAKNSLHEFSV-A 935
QY 382 YPQGLVDVTTITSEGLQGYKSNQVDFYNIPLAFRTSBDGLRRE-----427
DB 936 DEQGN--NFTVSNP---YSSYDTSKTSN-----VITPAGENGITTKVNGVVRVGDQTK 985
QY 428 -----GNHYSATLDSGGPEFGLPQTSN-----YKLSVNOLETSNV---DMS 467
DB 986 GLTTPKLTGVNNGKGVIVDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDEK 1045
QY 468 REMVNMIIQGFQMNKS-----VTTADTM 493
DB 1046 RAASIVDLVLSAGFNLOQNGEAVDFVSTYDTV 1076
```

[illegible]

Db 1444 GDTGT-----TAKKLGETLTI--KGGQTDNTKLTNNIGVAGTDGFTVKLAKDLTNL 1494
QY 106 HYTRAGNFRFTQDGF-----LNDP-----SGFTLMGSRISNNPNIKKETTLEPIQLD 151
Db 1495 NSVNAGGTRIIDEKGISFVDANGQAKANTFVLSANGLDLGGKRISN-----IGAAYD 1545
QY 152 FNDPTVAK---SPAKTSTALNAVNLGSDTDKTOSEANPYFALLESMKGNGTPISTSNY 208
Db 1546 DNDVNFQFNEVAKTVNNANQNSGASLPFVVDAN-----GKPINGTDG 1592
QY 209 SYAQPMRVYD---QQGNSHDITVFDGAPSTGSKTFEYLVAMNPSBDGS---AASGTD 262
Db 1593 KPQKAIKGADGKYHANANGVFDKDGKPIITDADKLANLAHAGKPLDAGHQVWASLGNS 1652
QY 263 AGLLMSGTM-----TFSSN-GEUKMTAFTPTGSATKDLNAPQAPLVNGLPQFS 311
Db 1653 DAITLTNKTSLPQLDTPNTGNAGQOQL-----PLSAAQQSNAASVKDVLN----- 1702
QY 312 ANFVGAGIQP--LTLDGFIKSOQ--NMWAGAPASAAAIGTDIGKLPSPMPIOTS----- 361
Db 1703 ---VGFNLQTNHNQVDF-VKAYDTVNFVNGTGADITSVRSADGTM-SNITVNTALAATDD 1757
QY 362 SGN--STARNGSSSTRYSODG-YEQGLVDVDTITSEKL-----QKYSNSQVY 408
Db 1758 DGNVLIKAKOG---KFKYADDLMPNGSLKAGKASDAKPTPTGLSLVNPAGKSGTGDV 1813
QY 409 DFYNIPLARF-----TSEDGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNQ 458
Db 1814 ALNLSKAVPKSKDGTITTTTSSDGI SIQKDNSSITLSKDGLNVGKVISNVGK-GTKD 1872
QY 459 LETSNVMSREMVMNI 474
Db 1873 TDAANVOQLNEVRNLL 1888

Search completed: October 26, 2004, 09:09:30
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2004, 03:22:21 ; Search time 477 Seconds
(without alignments)
5524.548 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586
Sequence: 1 MMGSFIGATGKMTHTGLG.....NSKSVTTADTQLKALELKR 502

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US1009823/runat_26102004_100128_1596/app.query.fasta_1.647
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US1009823@cgn_11_352/runat_26102004_100128_1596 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002as: *
7: Geneseqn2002bs: *
8: Geneseqn2003as: *
9: Geneseqn2003bs: *
10: Geneseqn2003cs: *
11: Geneseqn2003ds: *
12: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2586	100.0	1509	4 AAC88037	Aac88037 Lawsonia
2	577	22.3	568	2 AAT69214	Aat69214 Lawsonia
3	570	22.0	569	2 AAT69204	Aat69204 Lawsonia
4	529.5	20.5	1242	8 ACA24087	Aca24087 Prokaryot
5	528.5	20.4	1416	11 ABD04351	Abd04351 Pseudomon
6	528.5	20.4	1512	11 ABD04124	Abd04124 Pseudomon

7	527.5	20.4	1389	8 ABT14574	Abt14574 Pseudomon
8	509	19.7	110000	2 AAX20248_06	Continuation (7 of
9	509	19.7	111309	2 AAX20250	Aax20250 Borrelia
10	506.5	19.6	1239	8 ACA26709	Aca26709 Prokaryot
11	491.5	19.0	1209	12 ADK13720	Adk13720 E. coli
12	484.5	18.7	110000	6 ABQ69245_07	Continuation (8 of
13	484.5	18.7	110000	6 ABQ67197_06	Continuation (7 of
14	484.5	18.7	110000	6 ABAQ3041_07	Continuation (8 of
15	483	18.7	2550	2 AAV99813	Aav99813 DNA encod
16	476	18.4	10461	2 AAX20553	Aax20553 Polynucle
17	475.5	18.4	1239	10 ADF03725	Adf03725 Bacterial
18	469.5	18.2	5998	6 ABQ71042	Abq71042 Listeria
19	448.5	17.3	1215	10 AC668483	Acf668483 Photorhab
20	448.5	17.3	102644	10 ACF65378	Acf65378 Photorhab
21	448.5	17.3	110000	10 ACF67367_12	Continuation (13 of
22	414.5	16.0	110000	6 ABA92787_3	Continuation (4 of
23	360.5	13.9	1912	5 AAS88896	Aas88896 DNA encod
24	358.5	13.9	110000	2 AAX20248_00	Aax20248 Borrelia
25	307	11.9	810	11 ABD04540	Abd04540 Pseudomon
26	302	11.7	1545	11 ABD04219	Abd04219 Pseudomon
27	295	11.4	783	10 ACF68481	Acf68481 Photorhab
28	291	11.3	32768	2 AAX20515	Aax20515 Polynucle
29	288	11.1	816	2 AAT67783	Aat67783 H. pylori
30	288	11.1	837	2 AAT77463	Aat77463 H. pylori
31	288	11.1	837	2 AAT68116	Aat68116 H. pylori
32	286	11.1	798	10 ADF03653	Adf03653 Bacterial
33	281.5	10.9	1800	3 AAZ88552	Aaz88552 C. jejuni
34	279	10.8	1398	5 AAS88418	Aas88418 DNA encod
35	277.5	10.7	1800	2 AAV58977	Aav58977 FlgF ope
36	237.5	9.2	789	3 AAZ88554	Aaz88554 C. jejuni
37	237	9.2	555	6 ABK74597	Abk74597 Bacillus
38	221	8.5	742	12 ADM91627	Adm91627 Lawsonia
39	215.5	8.3	1524	6 ABK74555	Abk74555 Bacillus
40	201	7.8	1432	12 ADG32119	Adg32119 DNA encod
41	193	7.5	5445	4 AAI97965	Aai97965 Lawsonia
42	193	7.5	5445	9 ACA92323	Aca92323 Lawsonia
43	193	7.5	5445	10 ADG33882	Adg33882 L. intrac
44	193	7.5	5445	10 ADJ66787	Adj66787 Lawsonia
45	176	6.8	3129	11 ABD15269	Abd15269 Pseudomon

ALIGNMENTS

RESULT 1
AAC88037
ID AAC88037 standard; DNA; 1509 BP.
XX
AC AAC88037;
XX
DT 07-MAR-2001 (first entry)
XX
DE Lawsonia intracellularis flagellar hook protein FlgE gene SEQ ID NO:2.
XX
KW Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;
KW intestinal disease; immunogenic; diagnosis; antibacterial; swine; pig;
KW infection; detection; identification; ds.
XX
OS Lawsonia intracellularis.
XX
PN WO200069904-A1.
XX
PD 23-NOV-2000.
XX
PF 11-MAY-2000; 2000WO-AU000437.
XX
PR 13-MAY-1999; 99US-0133973P.
XX
PA (PTIZ) PFIZER PROD INC.
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.
PA (PIGR-) PTG RES & DEV CORP.
PI Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;
PI Ankenbauer RG;

XX WPI; 2001-016210/02.
 DR P-PSDB; AAB36552.
 XX
 PT New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody,
 PT useful in vaccines and diagnosis of Lawsonia infections, particularly in
 PT swine.
 XX
 PS Claim 32; Page 90-94; 97pp; English.
 XX
 CC The present sequence encodes the Lawsonia intracellularis flagellar hook
 CC protein FlgE. The present invention describes an isolated or recombinant
 CC polypeptide (I) that comprises, mimics or cross-reacts with a B- or T-
 CC cell epitope of a FlgE (flagellar hook) polypeptide from a Lawsonia spp.
 CC (I) has antibacterial activity, and induces a specific humoral immune
 CC response. (I) are used as antigens in vaccines to prevent or treat
 CC infection by Lawsonia, in birds and animals, especially pigs, to raise
 CC specific antibodies (Ab) and to detect past or present infection. Ab are
 CC also useful in diagnosis, to detect L. intracellularis or immunologically
 CC cross-reactive species, also for identification of epitopes in FlgE.
 CC Vectors that contain nucleic acids (II) encoding (I) are also useful in
 CC genetic vaccines, and fragments of (II) are useful as primers or probes
 CC for detecting L. intracellularis or related microorganisms, in
 CC hybridisation or amplification assays
 XX
 SQ Sequence 1509 BP; 511 A; 264 C; 316 G; 418 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3 88e-199 Length: 1509
 Score: 2586.00 Matches: 502
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-009-823A-1 (1-502) x AAC8037 (1-1509)

QY 1 MetMetGlySerLeuPheHeIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 1 ATGATGGGGAGTTGTTTATGTGTGCAACAGGTATGAAACCCATAGTACAGGGTGGGT 60
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 61 ACTGTCTCCATATATTCCTAACGCAATACATTTGGGTATAGCAGCAACAGGTAGTG 120
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 DB 121 TTTCAAGACCTGTTTATGTCAAGATTTAGCAATAGTCTTACTGGAGTCAAGGGGCAAC 180
 QY 61 GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPhe 80
 DB 181 CAGGCTGGTATGGGACACAGGTTGGAGTGTTCGCACAAATTTTACACAGGTCCTTT 240
 QY 81 GluProGlyAsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGlnValThr 100
 DB 241 GAACCTGGCAATAGTGTAAACAGATCTTGTCTATTTGGTGGAAAGGTTTTTTTCAGGTTACA 300
 QY 101 LeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPhe 120
 DB 301 TTACAGATTAAGTACATATACACAGCGAGGAATTTTCGTTTACTCAGATGGTTTT 360
 QY 121 LeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIle 140
 DB 361 TTAATGATCCTAGCGATTTACTTTAATGGGCTCAAGAATATCTAATAATCCTAACATA 420
 QY 141 LysLysGluThrLeuGluProIleGlnLeuAsnPheAsnAspProThrValAlaLysSer 160
 DB 421 AAAAGGAACCCCTTGAACCAATTCAGTTAGATCTTTAATGATCCTACAGTAGCAAGTCT 480
 QY 161 ProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLys 180
 DB 481 CCTGCAAAACAGTACAGCATTAACCGCTGTGGTAAACCTTGGTGTAGTAGTACAGATAAA 540

181 ThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr 200
 DB 541 ACACAAAGTGAAGCTAATCCACTACTTTCGAGAGCTGGAAAGGAAATGAACA 600
 QY 201 ProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln 220
 DB 601 CCTCCTATTTCTCATCAAACTACTCATATGCAACCTATGAGATGATATGATCAACAA 660
 QY 221 GlyAsnSerHisAspIleThrValTyrPheAspGlyValaProSerSerThrGlySerLys 240
 DB 661 GGAATTTCTCAGATATPACTGTATTTTGTAGTGAGCACCTCTTCAACAGGAGTAA 720
 QY 241 ThrPheGluTyrIleuValAlaMetAsnProSerGluAspGlySerAlaIleSerGlyThr 260
 DB 721 ACATTTCAATACCTTGTAGCTATCAATCCTAGTGAAGATGGAAGTGTGTCATCAGGAACA 780
 QY 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 DB 781 GATAGTCAGGTCTCTTAATGTCCTGNACTATGACATTTTCAAGTAAATGGCGAATTAATA 840
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaIleProGlnPro 300
 DB 841 AATATGACAGCTTTTACTCTCTCTGCTCTGCAACAAAGATTTAAATGTCATGGCAACCA 900
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 DB 901 GCACCATAGTCAATGGTTTACCACACAGTTTTCAGCAAAATTTTGTGTGCGGAATACAG 960
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
 DB 961 CCTTTAATACCTAGACTTTTGGAAATTAAAGGCCAACAGATATGTGGCAGGAGTCCAGCA 1020
 QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 DB 1021 TCCGCTCTGCCATAGGTACAGATATTTGGGAATTTGCCAATGATGCCAATACAAACA 1080
 QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
 DB 1081 TCCAGCGGTAAATCTACAGCAAGAAATGGATCATCTTCAACAGAGATATAGCCAAGAT 1140
 QY 381 GlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGly 400
 DB 1141 GGTTATCCTCAGGAGATCTAGTAGATGTCAATATACCTCTCGAAGGGAATTTACAAGGT 1200
 QY 401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420
 DB 1201 AAGTATAGTAAATAGTCAGGTGTGTTGATTTTATATATATTCCTTTAGCAGCTTTACAAGT 1260
 QY 421 GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440
 DB 1261 GAGGATGATTAAGACAGAGAGGGAATAACCATTTATTCGCCAACACTTGACTCAGGTGGG 1320
 QY 441 ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
 DB 1321 CCAGAGTTTGGATTCGCCAGAACATCTACTATCGGAAACTTAGTGTGTAATCACTTAG 1380
 QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe 480
 DB 1381 ACTTCTAACCTAGACATGACAGAGAAATGGTTAATATGATTTATTTCAACGCTGTTTT 1440
 QY 481 GlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500
 DB 1441 CAGATGAATAGTAAATCTGTTCACAGCAGACACAATGCTACAAAAAGCACTTGAACATA 1500
 QY 501 LysArg 502
 DB 1501 AAGCGT 1506

RESULT 2

AAT69214/c

ID AAT69214 standard; DNA; 568 BP.

XX

AC AAT69214;

XX DT 20-AUG-1997 (first entry)
XX DE Lawsonia intracellularis vaccine candidate DNA.
XX KW Intestinal disease; porcine proliferative enteropathy; vaccine; ss.
XX OS Lawsonia intracellularis.
XX PA WO9720050-A1.
XX PR 05-JUN-1997.
XX PR 29-NOV-1996; 96WO-AU000767.
XX PR 30-NOV-1995; 95AU-00006910.
XX PR 30-NOV-1995; 95AU-00006911.
XX PA (DARA-) DARATECH PTY LTD.
XX PA (PIGR-) PIG RES & DEV CORP.
XX PI Panaccio M, Hasse D;
XX WPI; 1997-310605/28.
XX Vaccine for treating or preventing Lawsonia intracellularis infection -
PT especially in pigs, containing non-pathogenic form of bacterium or its
PT components.
XX Claim 25; Page 71; 94pp; English.
XX DNA molecules (AAT69203-15) were isolated from an immunoscreening of a
CC Lawsonia intracellularis library using experimental sera from vaccinated
CC pigs. These nucleic acids, as well as isolated GroEL and GroES (AAT69201-
CC 02) sequences, encode putative vaccine candidates (see also AAM16678-85)
CC useful for protection of animals and birds against intestinal diseases,
CC esp. protection of pigs against porcine proliferative enteropathy (PPE).
CC They can also be used as genetic vaccines.
XX SQ Sequence 568 BP; 184 A; 126 C; 72 G; 176 T; 0 U; 10 Other;
Alignment Scores:
Pred. No.: 1,42e-37 Length: 568
Score: 577.00 Matches: 119
Percent Similarity: 95.97% Conservative: 0
Best Local Similarity: 95.97% Mismatches: 4
Query Match: 22.31% Indels: 2
DB: 2 Gaps: 0

US-10-009-823A-1 (1-502) x AAT69214 (1-568)
QY 1 MetMetGlySerLeuPheIleGlyAla-ThrGlyValSerThrHisSerThrGlyLeuG1 20
DB 370 ATGATGGGGAGTTTGTATTGTTGTCACACAGGTATGAAACCCATAGNACAGGNT-GG 312
QY 20 yThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVa 40
DB 311 TACTGTTCTCAATATATTGCTTACGCCAATACCATTTGGGTATAGCAGCACACAGTAGT 252
QY 40 lPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAs 60
DB 251 GTTTCAGACCTGTTTAGTCAAGATTAGCAATAGCTTTTACTGGAAGTCAGGGGCCAAA 192
QY 60 nGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPh 80
DB 191 CCAGGCTGGTATGGGAGCAGAGTTGGAGTGTTCACAAATTTTACACAGGGTGCCTTT 132
QY 80 eGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePheGlnValTh 100
DB 131 TGAACCTGGCAATAGTGTACAGATCTTCTATTGTGGGAAAAGGTTTTTTTCAGGTAC 72
QY 100 rLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPh 120

DB 71 ATTAGAGTAAAGTACTACTATACAGCAGGGAATTTTCGTTTACTCAAGATCGTTT 12
QY 120 eLeuAsnAsp 123
DB 11 TTTAAATGAT 2
RESULT 3
AAT69204
ID AAT69204 standard; DNA; 569 BP.
XX AC AAT69204;
XX DT 20-AUG-1997 (first entry)
XX DE Lawsonia intracellularis vaccine candidate DNA.
XX KW Intestinal disease; porcine proliferative enteropathy; vaccine;
XX KW flagellar basal body rod protein; ss.
XX OS Lawsonia intracellularis.
XX FH Key Location/Qualifiers
XX CDS 209..568
XX FT /*tag= a
XX FT /note= "includes in-frame stop codon at 245..247"
XX W09720050-A1.
XX 05-JUN-1997.
XX 29-NOV-1996; 96WO-AU000767.
XX PR 30-NOV-1995; 95AU-00006910.
XX PR 30-NOV-1995; 95AU-00006911.
XX (DARA-) DARATECH PTY LTD.
XX PA (PIGR-) PIG RES & DEV CORP.
XX PI Panaccio M, Hasse D;
XX WPI; 1997-310605/28.
XX P-PSDB; AAM16680.
XX Vaccine for treating or preventing Lawsonia intracellularis infection -
PT especially in pigs, containing non-pathogenic form of bacterium or its
PT components.
XX Claim 15; Page 51-52; 94pp; English.
XX DNA molecules (AAT69203-15) were isolated from an immunoscreening of a
CC Lawsonia intracellularis library using experimental sera from vaccinated
CC pigs. These nucleic acids, as well as isolated GroEL and GroES (AAT69201-
CC 02) sequences, encode putative vaccine candidates (see also AAM16678-85)
CC useful for protection of animals and birds against intestinal diseases,
CC esp. protection of pigs against porcine proliferative enteropathy (PPE).
CC They can also be used as genetic vaccines. The polypeptide (AAM16680)
CC encoded by the DNA molecule given in AAT69204 has sequence similarity to
CC flagellar basal body rod protein
XX SQ Sequence 569 BP; 177 A; 72 C; 128 G; 189 T; 0 U; 3 Other;
Alignment Scores:
Pred. No.: 5.23e-37 Length: 569
Score: 570.00 Matches: 118
Percent Similarity: 95.16% Conservative: 0
Best Local Similarity: 95.16% Mismatches: 5
Query Match: 22.04% Indels: 2
DB: 2 Gaps: 0

US-10-009-823A-1 (1-502) x AAT69204 (1-569)
QY 1 MetMetGlySerLeuPheIleGlyAla-ThrGlyMetLysThrHisSerThrGlyLeuG1 20

Db 200 ATGATGGGAGTTGTTTATTGTTGGAACAGGTATGAAACCCATAGNACAGGGNT-GG 258
Qy 20 YThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVa 40
Db 259 TACTGTCTCCAATAATATTCTTAACGCAATACCAATGGGTATAGCAGCAACAGGTACT 318
Qy 40 LpheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAs 60
Db 319 GTTTCAGACCTGTTTGTAGTCAAGATTTAGCAATAGGTTTACTGGAAGTCAGGGGCCAAA 378
Qy 60 nGlnAlaGlyMetGlyValGlnValGlySerValArgThrIlePheThrGlnGlyAlaPh 80
Db 379 CCAGGCTGTATGGAGCACAGGTGGAGGTGTCGACAAATTTTACACAGGGTGCTTT 438
Qy 80 eGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValTh 100
Db 439 TGAACCTGGCAATAGTGTAAACAGATCTGCTATTTGGTGGAAAGGTTTTTTTCAGGTTAC 498
Qy 100 rLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPh 120
Db 499 ATTAGAGCATTAAGTACACTATACAGCAGCGGGAATTTTCGTTTACTCAAGTGGTTT 558
Qy 120 eLeuAsnAsp 123
Db 559 TTTAAATGAT 568

RESULT 4

ACA24087 standard; DNA; 1242 BP.

AC ACA24087;
XX AC

DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #5744.
XXXX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.

XX OS Borrelia cepacia.

XX XN W0200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;XX DR WP1; 2003-029926/02.
XX DR P-PSDB; ABU20217.XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 14; SEQ ID NO 11957; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: the sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 1242 BP; 266 A; 421 C; 397 G; 158 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,57e-33	Length:	1242
Score:	529.50	Matches:	149
Percent Similarity:	44.20%	Conservative:	72
Best Local Similarity:	29.80%	Mismatches:	182
Query Match:	20.48%	Indels:	97
DB:	9	Gaps:	13

US-10-009-823A-1 (1-502) x ACA24087 (1-1242)

Qy	8	GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla	27
Db	16	GGCTTGAGCGGATGGCGGGTGGTGGAGCAATCTCGAGCATCTCGACATCGCAACATCGCG	75
Qy	28	AsnAlaAsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGln	47
Db	76	AACGAAACACGGTGGCTTCAAGCAGGAGCGGCAACTTCGCCGACATGTACGGGAAT	135
Qy	48	AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln	67
Db	136	TCGGTCGCGAGCTCGGTCAAC-----ACGAGATCGGCATCGCAGCGG	180
Qy	68	ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr	87
Db	181	CTCGCGTGGTGCAGCAGAAATTCGGCCAGGAGCATCAATTCGACGAGTCTCGCTC	240
Qy	88	AspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyr	107
Db	241	GACTTCGCGATCAACGGCAACGGCTTCTTCCAGATGTCCAGCAACGGCGTACCGGTAC	300
Qy	108	ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe	127
Db	301	TCGGCGCAGCGCAGTTCCTCATCGCACAGAACGGCGCATCTCGCAGCGCAAGCGCGC	360
Qy	128	ThrIleuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluPro	147
Db	361	AACCTGATGGGCTATGCGCGCGCGCGCGCGGTGTATCAACACCGCGCAGACCGTCCG	420
Qy	148	IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla	167
Db	421	CTCCAGCGC-----CCGACCAACAACATCGCGCGCGCGCGCAGCAGC---AAG	465
Qy	168	LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro	187
Db	466	ATCACCGCGCGCAGTTCACCTG---AACGGCGCAGCAAGGTCCGCGCCAG-----	513

QY 188 TyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProProIleSerThrSerAen 207
Db 514 -----ACGCGGTTCACCGCAGCGACAC 537
QY 208 -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp 225
Db 538 ACACGCTACAACTACACTCGTCGTCAGGCTACGACACGCTCGCGGCTCGACGAG 597
QY 226 IleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeu 245
Db 598 GTCCAGATGTACTTC-----GCGAAGAGCGCGCGCGCAGC----- 633
QY 246 ValAlaMetAsnProSerGluAspGlySerAlaAsaSerGlyThrAspSerAlaGlyLeu 265
Db 634 -----TGCTGGCCCTACGCGGCGGTGTCAGGCGCAGACGCGCG 669
QY 266 LeuMetSerGlyThrMetPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPhe 285
Db 670 ACGAATCTCGGCACCGCTCGATCGTCGTCGCGCGGATGATCGTCGACGCTGCTG--- 726
QY 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsn 305
Db 727 GCGCGACCGCGC-----CAGCCGACGCGC----- 750
QY 306 GlyLeuProGlnPheSerAlaAsnPheVal-----GlyAlaGlyIleGlnPro 321
Db 751 AGCCTCGGCAGTTCGCTTCGATCCGACACGACGCGCGCGCCCAATCGCAGAAC 810
QY 322 LeuThrIleuAspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProLaser 341
Db 811 CTGACGCTCGACCTG----- 825
QY 342 AlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSer 361
Db 825 ----- 825
QY 362 SerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGly 381
Db 826 ---GGCGACGACGACGAGTACGCGCGCAAGGACGCGGTGACCAATCTCGCGACGACGCG 882
QY 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 883 TTGCGAGCGGCGACGCTGACGAGGTCTCGATCGGCACGCGCGCAAGTGTACCGCGCAAC 942
QY 402 TyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGlu 421
Db 943 TACTCGAACCGCCAGACGCGGTGCTCGCGCTCGCAACTTCACCAACCGC 1002
QY 422 AspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyPro 441
Db 1003 AACGGGCTCGTGAACATCGCGCGCAACCAAGTATCGGAAACGCGCGCGTCCGCGCGCG 1062
QY 442 GluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThr 461
Db 1063 CAGATCGCGCGCGCGGACGACGACCAACCGACGCTCGCGGAGCGCGTGGAAAT 1122
QY 462 SerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGln 481
Db 1123 TCGAAGCTGAACCTGACGACCGAGCTCGTGAACCTGATCAGCGCGCGCGCAATACCA 1182
QY 482 MetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 1183 GCGAAGCGGACGATCAAGACCCAGCGCGGTGTCACGACGCTGTGTAACCTGCT 1242

RESULT 5
ABD04351
ID ABD04351 standard; DNA; 1416 BP.
XX AC ABD04351;
XX AC ABD04351;
XX AC ABD04351;
DT 29-JUL-2004 (first entry)
XX

DE Pseudomonas aeruginosa polynucleotide #2955.
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX Pseudomonas aeruginosa.
XX US6551795-B1.
XX 22-APR-2003.
XX 18-FEB-1999; 99US-00252991.
XX 18-FEB-1998; 98US-0074788P.
XX 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
XX P-PSDB; ABO70780.
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX Disclosure; SEQ ID NO 2955; 455pp; English.
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using biochip technology. Sequences ABD01397-
XX ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USFTO at
XX seqdata.uspto.gov/sequence.html
SQ Sequence 1416 BP; 295 A; 509 C; 407 G; 205 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,65e-33 Length: 1416
Score: 528.50 Matches: 151
Percent Similarity: 45.84% Conservative: 86
Best Local Similarity: 29.21% Mismatches: 205
Query Match: 20.44% Indels: 75
DB: 11 Gaps: 16

US-10-009-823A-1 (1-502) x ABD04351 (1-1416)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 31 AGTTTCAACATCGCGCTCGACGCGCATCCAGCGCGCTCTAGCGGCTTGAACGTCACGCGC 90
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGluAsp 43
Db 91 AACAACTATCGCAACGCGCGCATCGGTAGGCTTCAAGCAATCCCGCGGAGTTCCGCGAC 150
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db 151 GTCTACGCGCGCTCGGTG---CTGGGTTCG-----GGCAGCAACCGCAGGCG 195
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 196 AGCGCGGTGTGCTCTCGGACGCTCTCGAGATGTTCAAGCAGGCGGCAACATCGACTCGACC 255

QY 84 AsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGlnValThrLeuGluAsp 103
DB 256 AACAGCGTGTGGACCTGGCCATCAACGGCAACGGCTCTTCGTCAACGACCAACAGCGG 315
QY 104 LysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
DB 316 GCGATCAGCTACACCGCGCGGCTACTTCAATACCGAACAGCAGGATTTTCATCGTCGAC 375
QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
DB 376 AACACGGCTACCGCTGCAGGGCTATGCGCTCGGCGCGAACGCCAGTTCGACAGCGG 435
QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 152
DB 436 GTGGTCACCGACCTCAAGGTCGAGCGCGCAATCAGGCGCGCAGGCCACCTCGAGCATC 495
QY 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182
DB 496 CAGCAGTCGTCAACCTCACTCGACGCTGAG-----CCG 531
QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProPro 202
DB 532 CCGACCGTGCAGCGGTTTC-----GATCCG 555
QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsn 222
DB 556 TCGACGCGCGTACCTACAACCTCGTCTCTTCGTGGGCATCTATGACGCCAGGGGCNAC 615
QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
DB 616 TCCACACCACTAGCAGGATCTTCTCATCAAGAACGAGCGCGACCGCAATCGACCCCGCGCG 675
QY 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGlyAsp 253
DB 676 ATTCCGGAGAACAGCTGCACATGAAGTCTCATCGACGCGCTCATCGCTCGATCCG 735
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
DB 736 TCGAACCAAGACGGCG-----ATGAGCTTCAACGTCACCTTC 771
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
DB 772 GAGCCGACGGCCAGATGACCTCGTTCGGGCG-----CCGACGGCAGCACCAGCGGG 825
QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304
DB 826 CCGGGCTTCAGCATCGACGGCGACCCACACGTGATCCAGTTCAGTCCGGCC----- 876
QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
DB 877 ACGGGCATCCGGGACTCCC-----GGCACCGGCTGGATTCCGGCGGGCTCG 924
QY 325 AspPheGlyIleLysSerGlnAsnMetTrpAlaGlyAlaProLaserAlaAla 344
DB 925 GAC---GGCAAGACCCCGCGACCTACGCTGGAATGGCGCGACCGGTGCGCGACGGCG 981
QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
DB 982 ATTCTCTCGACATGGCCAG-----ACCACCCAGTAC 1014
QY 365 SerThrAlaArgAsnGlySerSerSerThrArgTyrSerGlnAspGlyTyrProGln 384
DB 1015 TCCACCGCGTTCGCCACGAGCAACCCGATC-----CAGGACGGCTACACACC 1062
QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
DB 1063 GGTACGCTGCAGGCGCTGGAATACGACACCGGGGTGATCTTCGCCCGCTACACCAT 1122
QY 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
DB 1123 GGCACGATCCAAAGTGCAGGCGCAGGTGGTGGTCCGCAACTCCGCAACATCCACGAGCGCTG 1182

Qy	425	ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly	444
Db	1183	ACGCCGATCGCAGAACCTCTCGGTGCATCGCGCGAGTGTCGGCAGCGCGGTCGGC	1242
Qy	445	LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal	464
Db	1243	CGCCGCGCTCGGCACCCTCGGGGGCGTTGCATCCGGCGCGCTGGGAAGCGTCCAACGTG	1302
Qy	465	AspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSer	484
Db	1303	GATCATCTCCAAACAAGACTGGTGAACCTCATCTGCCACGAGCGCACTACCGAGGCAATGCC	1362
Qy	485	LysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys	501
Db	1363	AAGACCATCAGACCGAGGATGCGGTGACCCAGACCATCATCACTGGCGC	1413
 RESULT 6 ABD04124/c ID: ABD04124 standard; DNA; 1512 BP.			
XX	AC	ABD04124;	
XX	DT	29-JUL-2004 (first entry)	
XX	DE	Pseudomonas aeruginosa polynucleotide #2728.	
XX	KW	Bacterial infection; Gene; ds; Pseudomonas aeruginosa infection;	
XX	KW	antibacterial.	
XX	OS	Pseudomonas aeruginosa.	
XX	PN	US6551795-B1.	
XX	PD	22-APR-2003.	
XX	PF	18-FEB-1999; 99US-00252991.	
XX	PR	18-FEB-1998; 98US-0074788P.	
XX	PR	27-JUL-1998; 98US-0094190P.	
XX	FA	(GENO-) GENOME THERAPEUTICS CORP.	
XX	PI	Rubenfield MJ, Nolling J, Deloughery C, Bush D;	
XX	DR	WPI; 2003-615309/58.	
XX	DR	P-PSDB; ABO70553.	
XX	PT	Novel isolated nucleic acid encoding pseudomonas aeruginosa polypeptide,	
XX	PT	useful as molecular targets for diagnostics, prophylaxis and treatment of	
XX	PT	pathological conditions resulting from bacterial infection.	
XX	PS	Disclosure; SEQ ID NO 2728; 455pp; English.	
XX	CC	The invention relates to pseudomonas aeruginosa polypeptides and the	
XX	CC	polynucleotides encoding them. The sequences are useful in diagnosis and	
XX	CC	therapy of pathological conditions, as molecular targets for diagnostics,	
XX	CC	prophylaxis and treatment of pathological conditions resulting from a	
XX	CC	bacterial infection, for evaluating a compound, such as a polypeptide,	
XX	CC	for the ability to bind a P. aeruginosa nucleic acid, as components of	
XX	CC	effective antibacterial targets, as targets for antibacterial drugs,	
XX	CC	including anti-P. aeruginosa drugs, as templates for recombinant	
XX	CC	production of P. aeruginosa-derived peptides or polypeptides, as target	
XX	CC	components for diagnosis and/or treatment of P. aeruginosa-caused	
XX	CC	infection, and in detection of P. aeruginosa sequences or other sequences	
XX	CC	of Pseudomonas species using biochip technology. Sequences AB01397-	
XX	CC	AB013967 represent P. aeruginosa polynucleotides of the invention. Note:	
XX	CC	The sequence data for this patent did not form part of the printed	
XX	CC	specification but was obtained in electronic format from USFTO at	
XX	CC	seqdata.uspto.gov/sequence.html	
XX	NO	Sequence 1512 BP: 224 A; 434 C; 543 G; 311 T; 0 U; 0 Other;	

Pred. No.: 3,96e-33 Length: 1512
 Score: 528.50 Matches: 151
 Percent Similarity: 45.84% Conservative: 86
 Best Local Similarity: 29.21% Mismatches: 205
 Query Match: 20.44% Indels: 75
 DB: 11 Gaps: 16

US-10-009-823A-1 (1-502) x ABD04124 (1-1512)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 DB 1503 AGTTTCAACATCGGCTGAGCGATCCAGCGGCTCTAGCGGCTGAACGTCAACCGC 1444
 QY 24 AsnAsnIleAlaAsnAlaThrIleGlyLysGlnGlnValValPheGlnAsp 43
 DB 1443 AACAAATCGCAACCGCGACCGTAGGCTTCAAGCAATCCCGCGGAGTTCGCCGAC 1384
 QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 DB 1383 GTCTACGCGCTCGGTG---CTGGGTTG-----GGCAGCAACCGCAGGCG 1339
 QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
 DB 1338 AGCGGGCTGTCTCTCGGACGCTCTCGCAGATGTTCAAGCAGGCAACATCGACTCGACC 1279
 QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAsp 103
 DB 1278 AACAGGCTGTGACTGCGCATCAACGCAACGCGTCTTCTGTCACGCAACACGCG 1219
 QY 104 LysValHisThrArgAlaGlyAsnPheA-gPheThrGlnAspGlyPheLeuAsnAsp 123
 DB 1218 GCGATCAGCTACACCGCGCGCTACTTCAATACGCAACGAGGATTTTCATCGTCAC 1159
 QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlu 143
 DB 1158 AACACGGCTACCGCTCGAGGCTATGCGGCCGCAACGCGGAGTTGCAAGACGCG 1099
 QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
 DB 1098 GTGGTACCGCACTCAAGTCGAGCGCGCAATCAGCGCGCGCAGCCACCTCGAGCATC 1039
 QY 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182
 DB 1038 CAGCAGCTCGTCAACCTCAACTCGACGCTGAAG-----CCG 1003
 QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 202
 DB 1002 CCGACCGTCAGCGCTTC-----GATCGG 979
 QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsn 222
 DB 978 TCCGACGCGCTACCTACAACTGCTCTCTGCTGGGCACTATGACAGCGCGCAAC 919
 QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
 DB 918 TCCACACCATGAGCGAGTCTTTCATCAGACGAGCGGCGCAATCGGACCGCGCGG 859
 QY 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
 DB 858 ATTCCGGAGAACAGCTGGACCATGAAGTGTGATCGACGGCGTCAATCCGCTCGATCGG 799
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 798 TCGAACAAGACGCGC-----ATGAGCTTCAACGTCACTTC 763
 QY 274 SerSerAsnGlyLeuLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
 DB 762 GAGCCGAGCGCGAGTACCTCGTTCGGCG-----CCGACGCGGACACGAGCGG 709
 QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304
 DB 708 CCGGGCTTCAGCATCGACGCGACCAACGATGATCCAGTTCAGTTCGCGCGC----- 658

QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
 DB 657 ACCGCAATCGCGCACTCC-----GGCACCGGCTGATTCGCGGCGCTCG 610
 QY 325 AspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAa 344
 DB 609 GAC---GGCAAGACCCCGCGACCTACGCTGGAAATGGCGGCGCGGCGCGGCGGCGG 553
 QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
 DB 552 ATCTCTTCGACATCGCAAG-----ACCACCCAGTAC 520
 QY 365 SerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyrProGln 384
 DB 519 TCCACCGCTTCGCCCGCAGACACCCGATC-----CAGGACGCTACACCACT 472
 QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
 DB 471 GGTGAGCTGGCAGCGCTGGAATCGACGACACCGGGTGATCTTCGCCCGCTACACCAAT 412
 QY 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
 DB 411 GGCCAGTCCCAAGGTGCGAGGCGCGTGTGTGCGCAACTTCGCCAATCCAGGGCGCTG 352
 QY 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
 DB 351 ACGCCGATCGCAAGACCTCTCGGTGCTCGCGGAGTCCGCGAGCGCGGCGTGGC 292
 QY 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
 DB 291 GCGCGCGCTCGGCGACCCCTGGGGCGCTTGCATTCGCGCGCGCTGGGAAGCGTCCAACGTG 232
 QY 465 AspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSer 484
 DB 231 GACATCTCCACGACGCTGGTGAACCTCATGTCCACGCGCAACTACGAGGCAATGCC 172
 QY 485 LysSerValThrAlaAspThrMetLeuGlnLysAlaLeuLysLeuLys 501
 DB 171 AAGACCATCAGACCGAGGATGCGGTGACCCAGACCATCATCACTCGCGC 121

RESULT 7
 ID ABT14574 standard; DNA; 1389 BP.
 AC ABT14574;
 DT 27-FEB-2003 (first entry)
 DE Pseudomonas aeruginosa biofilm formation-related gene #16.
 KW Gene; ds; biofilm formation modulation; biofilm-associated disease;
 KW cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
 KW catheter-associated infection; medical device-associated infection.
 OS Pseudomonas aeruginosa.
 FN WO200285295-A2.
 PD 31-OCT-2002.
 PF 19-APR-2002; 2002WO-US012532.
 XX 20-APR-2001; 2001US-0285190P.
 PR 24-OCT-2001; 2001US-0344142P.
 XX (IOWA) UNIV IOWA RES FOUND.
 PA (HARD) HARVARD COLLEGE.
 XX Whiteley M, Bangera MG, Lory S, Greenberg EP;
 DR WPI; 2003-075601/07.
 DR P-PSDB; ABJ18752.

PT Identifying compound capable of modulating biofilm formation by
 PT bacteria/bacterial antibiotic resistance, useful for treatment of biofilm
 PT associated disease.

XX Claim 1; Page 70; 154pp; English.

CC The invention comprises a method for identifying a compound capable of
 CC modulating biofilm formation by bacteria. The method of the invention is
 CC useful for identifying a compound capable of modulating biofilm formation
 CC by bacteria or modulating bacterial antibiotic resistance. The method of
 CC the invention is also useful for diagnosing and treating a subject
 CC (especially an immunocompromised human) that is afflicted with a biofilm-
 CC associated disease or disorder, such as: cystic fibrosis; AIDS; middle
 CC ear infections; acne; periodontal disease; catheter-associated infections
 CC; and medical device-associated infections. The present DNA sequence
 CC represents a gene that is used in the invention

XX SQ Sequence 1389 BP; 291 A; 501 C; 396 G; 201 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.29e-33 Length: 1389
 Score: 527.50 Matches: 151
 Percent Similarity: 45.84% Conservative: 86
 Best Local Similarity: 29.21% Mismatches: 205
 Query Match: 20.40% Indels: 75
 DB: 8 Gaps: 16

US-10-009-823A-1 (1-502) x ABT14574 (1-1389)

Qy 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
 Db 4 AGTTTCAACATCGCGCTGACGGCATCCAGCGGCTCTAGCGGCTGAACGTCACCGGC 63
 Qy 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
 Db 64 AACACATCGCCACCGCGCACCGTAGGTTCAAGCAATCCGCGGAGTTCCGGAC 123
 Qy 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
 Db 124 GTCTACGCCGCCCTCGGTG---CTGGGTTCG-----GGCAGCAACCGCGAGGC 168
 Qy 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
 Db 169 AGCGGCTGTCTCTCGACGCTCTGCAGATTTCAAGAGGGCAACATCGACTCGACC 228
 Qy 84 AsnSerValThrAspLeuAlaIleGlyGlyPhePheGlnValThrLeuGluAsp 103
 Db 229 AACAGCGTGTGGACCTGGCCATCAACGGCAACGGCTTCTTCACACCAACCAACGGG 288
 Qy 104 LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
 Db 289 GCATAGCTACACCGCGCGGCTACTTCAATACCGACAGCAGGATTTATCGTCGAC 348
 Qy 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
 Db 349 AACACGGCTACCGCTCGAGGCTATGCGTCCGGCCGACCGCCAGTTCGAGAACGGC 408
 Qy 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
 Db 409 GTGGTACCGACCTCAAGTTCGAGCGCGCAATCAGGCGCGCCACCTCGAGCATC 468
 Qy 163 LysThrSerThrAlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGln 182
 Db 469 CAGCAGTCTGTACACTCACTCGACGCTGAG-----CCG 504
 Qy 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202
 Db 505 CCGACCGTGCACCGGCTTC-----GATCGG 528
 Qy 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsn 222
 Db 529 TCCGACCGCGCTACCTACACTCGTCTCTTCGCTGGCATCTATCAGACCGCAGGCAC 588

Qy 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
 Db 589 TCCACACCATGACCGAGTCTTCATCAAGACGCGCGGACCGAATGCGACCGCGCG 648
 Qy 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
 Db 649 ATTCCGAGAGACACTGGACCATGAAGTGTGATCGACGCGCTCAATCGCTCGATCCG 708
 Qy 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 Db 709 TCGAACAAGACCGCG-----ATGAGCTTCAACGTCACCTTC 744
 Qy 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
 Db 745 GACGCCAGCGCCAGATGACTCGTTCGGTTCGGCG-----CCGACGCGCAGCAGCGGG 798
 Qy 291 -----AlaThrLysAspLeuAsnAlaThrGlnProAlaProLeuVal 304
 Db 799 CCGGCTTCAGATCGACGCGGACCAACGATGATCCAGTTCAGTCCGGC----- 849
 Qy 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
 Db 850 ACTGCAATCGCGGACTCC-----GGCACCCTGCTGATTCGGCGGCGCTCG 897
 Qy 325 AspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 344
 Db 898 GAC---GGCAAGACCCCGCGCACTACGCTGGAACGCGCGACCGGTGCGCGCAGCGGC 954
 Qy 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
 Db 955 ATCTCTTCGACATCGCGCAG-----ACCACCCAGTAC 987
 Qy 365 SerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAspGlyTyrProGln 384
 Db 988 TCCACCGGTTCCGCCAGACCAACCGCATC-----CAGACGCTACACCAAC 1035
 Qy 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
 Db 1036 GTGATGTTGGCAGCGCTGGAAATCGACACACCGGGGTGATCTCCGCGCTACACCAAC 1095
 Qy 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
 Db 1096 GGCCAGTCCAAGTGCAGGCGCAGGTGCTGCGCAACTTCGCCAATCCACATCCAGGCGCTG 1155
 Qy 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
 Db 1156 ACGCCGATCGGCAAGACTCTGCTGGTGCAGTCTTCGGAGTCCGCGAGCGCGGCGTCCG 1215
 Qy 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
 Db 1216 GCGCGCGCTCGGCGCACCCCTGGGGCGGTTCGAATCCGCGCGCTGGAAGGCTCAACGTG 1275
 Qy 465 AspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSer 484
 Db 1276 GACATCTCCAACGAACTGGTGAACCTCATGTCACCGCGCAACTACACGCGCAATGCC 1335
 Qy 485 LysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 Db 1336 AAGACATCCAGACCGAGGATCGGTGACCCAGACCATCATCAACCTGCGC 1386

RESULT 8

AAAX20248_06
 Continuation (7 of 10) of AAAX20248 from base 600001 (Borrelia burgdorferi polynucleotide
 WP Sequence split into 10 fragments LOCUS AAAX20248 Accession Aax20248

WP	Fragment Name	Begin	End
WP	AAAX20248_00	1	110000
WP	AAAX20248_01	100001	210000
WP	AAAX20248_02	200001	310000
WP	AAAX20248_03	300001	410000
WP	AAAX20248_04	400001	510000
WP	AAAX20248_05	500001	610000
WP	AAAX20248_06	600001	710000
WP	AAAX20248_07	700001	810000

WP AAX20248_08 800001 910000
 WP AAX20248_09 900001 910715
 Alignment Scores:
 Pred. No.: 3,178-29 Length: 110000
 Score: 509.00 Matches: 150
 Percent Similarity: 41.62% Conservative: 71
 Best Local Similarity: 28.25% Mismatches: 190
 Query Match: 19.68% Indels: 120
 DB: 2 Gaps: 13

US-10-009-823A-1 (1-502) x AAX20248_06 (1-110000)
 QY 1 MetMetGlySerLeuPheLeuGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 15932 ATGATGAGCTCTTATATCTGCTCTTCTGCTCTTCTGAGATCATCAACCAAGATGGAT 15991
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 15992 GTTGTGGTAAACAATATCCCAATGTAATACAAATGCTTTAAAGGAGAGATAAT 16051
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 DB 16052 TTTCAAGATATGATCGCAATCTATTCTGAGCTTCTCGCCCTACTGATGCTCGTGGT 16111
 QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 DB 16112 GGGACTAATCCCAAGCAAGTTGGATTAGCATGAATGTCCTCAATGACACTATTAC 16171
 QY 76 ThrGlnGlyValaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
 DB 16172 ACTCAAGAGCTTTTCAAGACACTCAAAAGCATCTGATCTGGAGTTAGTGGCAACGGA 16231
 QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 DB 16232 TTTTATTTTAAAGAGGTAATAATTTGTTTTATACAGAGCGCGTCTTTGACGTG 16291
 QY 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer 135
 DB 16292 GACTCTGATCGACATCTTGTAAATCTGCAAT-----GGATCGCAATTCAA 16339
 QY 136 Asn-----AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 DB 16340 GGTGGATGGCAAGAGATTAGAGGTGAAAGGTTATAAATACAGCTTCTGATATTGAG 16399
 QY 154 Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
 DB 16400 GATCTGATTATCCGATTGGAGATAAAGAGGAGCAAGCTTACCAGAAAATGTTACTTTT 16459
 QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProThrPheAla 190
 DB 16460 GCTTGTAAATCTTGATAAGAGATTGCCCTTAATTCAGAGAGGTGGCAATCTGCAGATATT 16519
 QY 191 LeuLeuGluSerThrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyr 210
 DB 16520 GCACGCGGAACCTGGGTTCGTAATAATCA----- 16549
 QY 211 AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPhe 230
 DB 16550 -----TTGATATGACAGTTTGGAAAT-----GTTAGTCTTCT--- 16582
 QY 231 AspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnPro 250
 DB 16582 ----- 16582
 QY 251 SerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThr 270
 DB 16582 ----- 16582
 QY 271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
 DB 16583 -----GAGCTTAGA----- 16591

291 AlaThrLysAspLeuAsnAla-----TrpGlnProAlaProLeuValAsnGlyLeu 307
 DB 16592 GTTGTAAAGATTAAATACGCTAAATTTTGTGAATGCAACAGTATTATAAATGTT--- 16648
 QY 308 ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe--- 326
 DB 16649 ---GAGCAAAATTCAAATTTT-----ACACTGGGTTTGAC 16681
 QY 327 -----GlyIleLysSerGlnGlnAsnMetThrPalaGly 337
 DB 16682 AATGAAGGAGCATTTGGCTCTTTTAAATGGTCAACCAAGGCAAAAGAGATATTTCTTCAA 16741
 QY 338 AlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetPro 357
 DB 16742 ATTCCTATACATTTAATGTTTGGTGCMAATGATAGTGAAGTTGGTGGAGCAAACT 16801
 QY 358 IleGlnThrSerSerGly-----AsnSerThrAlaArgAsnGly 370
 DB 16802 GTAAATTTGAAATTTGGCAACAGTTGGAAGTTACACTGATTCATTAATCTCAGTTTGTCTGAT 16861
 QY 371 SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal 390
 DB 16862 TCTAGTAGCAACAAGGCTATTATTCAAGATGGATATGGCATGGGATATATGGGAAATAT 16921
 QY 391 ThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPhe 410
 DB 16922 GAATTTGATCAAAATGCTGTATTATGTTGGCATTTTCAATGGCATAGACGAGATCTT 16981
 QY 411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsn 430
 DB 16982 GCGAAGATTTCTCTGCTCTTTTATCAATCCCGAGGAGCTTCAAAATCAGCGCATCT 17041
 QY 431 HistySerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
 DB 17042 AATTTCTGAACAACAGCAATTCAGTCAAGTAGAATAGGCGAACTGGACTGCTGGA 17101
 QY 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
 DB 17102 CTGCTGATATTAGATCTGCTGTTTGTAGAAATGCCCAATGTTGATCTGCAGAGCAATTT 17161
 QY 471 ValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAla 490
 DB 17162 ACAGATATGATATGATGACCCCAAGAGATTTCAGGCAATGCCAAATGCCAATACCCTTCT 17221
 QY 491 AspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 DB 17222 GATCAATTATTACAGAACTTGTAAAGATTGAAA 17254

RESULT 9
 AAX20250
 ID AAX20250 standard; DNA; 111309 BP.
 AC AAX20250;
 DT 04-MAY-1999 (first entry)
 DE Borrelia burgdorferi polynucleotide sequence #3.
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.
 OS Borrelia burgdorferi.
 FN WO9858943-A1.
 PD 30-DEC-1998.
 XX 18-JUN-1998; 98WO-US012764.
 XX 20-JUN-1997; 97US-0050359P.
 PR 22-JUL-1997; 97US-0053344P.
 PR 22-JUL-1997; 97US-0053377P.

PR 03-SEP-1997; 97US-0057483P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Fraser C, White OR, Clayton R, Dougherty BA, Lathigra R;
 XX Smith HO;
 XX WPI; 1999-081217/07.
 DR
 XX
 XX New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention and
 PT therapy of infections, particularly Lyme disease.
 XX
 XX Claim 1; Page 738-800; 1128pp; English.
 PS
 CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
 CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for the
 CC detection, diagnosis, characterisation, prevention and therapy of Bb
 CC infections, e.g. Lyme disease. They can also be used for the production
 CC of biosynthetic products, e.g. enzymes. *Borrelia* belongs to a family of
 CC motile, spiral-shaped bacteria called Spirochetes. Spirochetes are
 CC pathogenic in humans and *Borrelia* causes epidemic and endemic relapsing
 CC fever, and Lyme borreliosis, more commonly known as Lyme disease
 CC
 SQ Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 0 U; 10 Other;

Alignment Scores:
 Pred. No.: 3,21e-29 Length: 111309
 Score: 509.00 Matches: 150
 Percent Similarity: 41.62% Conservative: 71
 Best Local Similarity: 28.25% Mismatches: 190
 Query Match: 19.68% Indels: 120
 DB: 2 Gaps: 13

US-10-009-823A-1 (1-502) x AAX20250 (1-111309)

QY 1 MetMetGlySerLeuPheHleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db
 QY 21 ThrValSerAsnAlaAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 Db
 QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 Db
 QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
 Db
 QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 Db
 QY 116 ThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIleSer 135
 Db
 QY 136 Asn-----AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 Db
 QY 154 Asp-----ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
 Db
 QY 27014 GATCTGATTATTCGGATTGGAGATAAGAGGGGAGCAAGTCTACGAAAAAATGTTACTTTT 27073

QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla 190
 Db
 QY 191 LeuLeuGluSerThrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyr 210
 Db
 QY 211 AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPhe 230
 Db
 QY 231 AspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnPro 250
 Db
 QY 251 SerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThr 270
 Db
 QY 271 MetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer 290
 Db
 QY 291 AlaThrLysAspLeuAsnAla-----TyrGlnProAlaProLeuValAsnGlyLeu 307
 Db
 QY 308 ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe--- 326
 Db
 QY 327 -----GlyIleLysSerGlnGlnAsnMetTyrAlaGly 337
 Db
 QY 338 AlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetPro 357
 Db
 QY 358 IleGlnThrSerSerGly-----AsnSerThrAlaArgAsnGly 370
 Db
 QY 371 SerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspVal 390
 Db
 QY 391 ThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPhe 410
 Db
 QY 411 TyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsn 430
 Db
 QY 431 HisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
 Db
 QY 451 TyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMet 470
 Db
 QY 471 ValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAla 490
 Db
 QY 491 AspThrMetLeuGlnLysAlaLeuGluLeuLys 501
 Db
 QY 27836 GATCAATTATTACAGAACTTGTAAAGATTGAAA 27868

ACA26709

ID ACA26709 standard; DNA; 1239 BP.
 AC ACA26709;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #8366.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 XX drug design; gene.
 XX
 OS Burkholderia mallei.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 PI
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU22839.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 14579; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1239 BP; 271 A; 424 C; 379 G; 165 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.84e-31 Length: 1239
 Score: 506.50 Matches: 137
 Percent Similarity: 43.34% Conservative: 81
 Best Local Similarity: 27.24% Mismatches: 180
 Query Match: 19.59% Indels: 105
 DB: Gaps: 11

US-10-009-823A-1 (1-502) x ACA26709 (1-1239)
 QY 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAlaAla 27
 DB 16 GGTGTGAGCGGTTGGCGGCGGTTCGAGCGACTCGAGTATCGGCACACATCGCG 75
 QY 28 AsnAlaAsnThrLeuGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGln 47
 DB 76 AACGCGAACACAGGTGGCGCTTCAAGGGCAGCACCGCGCAGTTTCCGACATGTATCGGAAT 135
 QY 48 AspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGln 67
 DB 136 TCGGTCCGA-----TCGGCGTCAACAATCCGATCGGCATCGGCACGATG 180
 QY 68 ValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThr 87
 DB 181 CTGCATCGGTGCAGCAGCAGTTCAGCCAGGCGACGATCACCTCGAGCAGCTCGCGCTG 240
 QY 88 AspLeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyr 107
 DB 241 AAGTCGCGATCAACGCGCTTCTCCAGATGTCGAACACCGCGGTGACACGATAC 300
 QY 108 ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe 127
 DB 301 TCGCGCGACGCGACGTTCCAGCGCGACAGACAGCGGTACATCGTCAATTCCGAGGCGCTG 360
 QY 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGluThrLeuGluPro 147
 DB 361 AACCTGATGGGTACGCGCGAACGCGCGGTGATCAACACCGCGCGCGCGCGCTG 420
 QY 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
 DB 421 CTGCAGCGG-----CCGACGACCAACATCGCGCCACCGCGACG---ACCAAG 465
 QY 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
 DB 466 ATCACCAGCGCAGTTCACCTG---AACTCGACGAGCGCGGTGCGCGCG---ACGACGCGG 519
 QY 188 TyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProLysSerThrSerAsn 207
 DB 520 TTC-----AACTACCGCATCGACGAC----- 543
 QY 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227
 DB 544 TACAACITACAGCAGCTCGGTGAGGTGTTCCGACACGCTCGCGGCTCGCAGAACGTGAAC 603
 QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
 DB 604 CTGTACTCTGTGAGAGCGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 663
 QY 248 MetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet 267
 DB 664 AAGACGCGGACCGCATCTCGGCTCG----- 687
 QY 268 SerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrPro 287
 DB 688 -----GTCAAGTTCAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714
 QY 288 ThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeu 307
 DB 715 ACGGGCAGCTCGACGCGC----- 732
 QY 308 ProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGly 327

Db 732 ----- 732
 Qy 328 IleLysSerGlnGlnAsnMetTTPAlaGlyAlaProAlaSerAlaAlaLalleGlyThr 347
 Db 733 -----GCGGCGTCCGACG-----ACG 750
 Qy 348 AspileGlyLysLeuProSerMetProileGlnThrSerSerGlyAsn----- 364
 Db 751 AACGTCGCCAGTCTCTGTTCTCGATCCCGACGACGCGGCGCGGACCGGAGAAC 810
 Qy 365 -----SerThAlaAraAsnGlySerSerSerThrArgArgTyr 377
 Db 811 CTGACGCTCGACCTCAGCGGACGAGCGAGTACGCGGCGCAAGAACCGCATCAACATCTC 870
 Qy 378 SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrileThrSerGluGlyLys 397
 Db 871 GCGCAGACGCGTTCGCGGCGGCTGCTGACGACGTTCTCGATCGCGGCGGACGCAAG 930
 Qy 398 LeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnileProLeuAlaArg 417
 Db 931 GTCACCGGCAACTACTCGAACGGCCAGACCTCCACGCTCGGCCAGATCGTGTCTCGCAAC 990
 Qy 418 PheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAsp 437
 Db 991 TTACACCCGACGCGCTCTGTGAACGTCGCGCAACACAGTACGTCTGAGTCCGCGCGG 1050
 Qy 438 SerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsn 457
 Db 1051 TCGGGGTGCGGAGATCTCCGCGCGCGGCGGACGACGCGGACGCGTGCAGGCGAC 1110
 Qy 458 GlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetilelleGln 477
 Db 1111 GCGCTCGAGAACTCGAACGTCGACCTCGACGAGCGACGCTGTGAAGCTGTATCATCCGCGCAG 1170
 Qy 478 ArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAla 497
 Db 1171 CGCACTACGCGGCGGAGCGGCGGACGACGACGACGACGACGACGACGACGACGCTC 1230
 Qy 498 LeuGluLeu 500
 Db 1231 ATCAACCTG 1239
 RESULT 11
 ADK13720
 ID ADK13720 standard; DNA; 1209 BP.
 XX AC ADK13720;
 XX 20-MAY-2004 (first entry)
 DT
 XX E. coli iron transport and metabolism protein encoding DNA SEQ ID NO:15.
 DE
 XX Escherichia coli; iron transport and metabolism protein; iron transport;
 KW iron metabolism; antibacterial; gene; ds.
 XX
 OS Escherichia coli.
 PH Location/Qualifiers
 FT CDS
 FT 1..1209
 FT /*tag= a
 FT /product= "iron transport and metabolism protein"
 PN WO2004018638-A2.
 XX
 PD 04-MAR-2004.
 XX
 PF 21-AUG-2003; 2003WO-US026488.
 XX
 PR 21-AUG-2002; 2002US-0405331P.
 XX
 PA (MINU) UNIV MINNESOTA.
 PA (KAPU/) KAPUR V.
 PA (GADG/) GADGIL M.

XX Kapur V, Gadgil M;
 XX WPI; 2004-238974/22.
 DR P-PSDB; ADK13784.
 XX
 PT New isolated and purified iron transport and metabolism polypeptides and
 PT encoding polynucleotides, useful in identifying potential targets for
 PT agents against pathogenic bacteria.
 XX
 PS Claim 31; SEQ ID NO 15; 185pp; English.
 XX
 CC The present sequence encodes an Escherichia coli iron transport and
 CC metabolism protein. Also described: (1) an isolated and purified
 CC polynucleotide comprising a nucleic acid sequence encoding an Escherichia
 CC coli iron transport and metabolism protein; and (2) an expression
 CC cassette comprising a nucleic acid sequence encoding a promoter operably
 CC linked to at least one of the polynucleotide sequences of (1). The
 CC Escherichia coli iron transport and metabolism proteins have
 CC antibacterial activity. The methods and compositions of the present
 CC invention are useful in identifying genes and proteins involved in
 CC bacterial iron transport and metabolism, and using such as potential
 CC targets for agents against pathogenic bacteria.
 XX
 SQ Sequence 1209 BP; 327 A; 320 C; 303 G; 259 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2.89e-30 Length: 1209
 Score: 491.50 Matches: 143
 Percent Similarity: 43.95% Conservative: 75
 Best Local Similarity: 28.83% Mismatches: 173
 Query Match: 19.01% Indels: 105
 DB: 12 Gaps: 15
 US-10-009-823A-1 (1-502) x ADK13720 (1-1209)
 Qy 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnileAlaAsnAla 29
 Db 22 AGCGATTAAACGCTGCGCGCCACCACTCGATGTTATTGGCAACATATCGCACTCC 81
 Qy 30 AsnThrileGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
 Db 82 GCCACTACGCTTTAAATCAGGCAGCGCTCTTTTCCCATATGTTGCG----- 132
 Qy 50 AlaileGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
 Db 133 -----GGTTCG-----AACTGGGACTGGGGTAAAGTTGCC 165
 Qy 70 SerValArgThrilePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
 Db 166 GGTATCACTCAGGACTTTACCGATGGCAGCACCAACACCGCGCGAGGCTCTGGACGTT 225
 Qy 90 AlaileGlyGlyLysGlyPheGlnGlnValThrleuGluAspLysValHisTyrThr 108
 Db 226 GCTATCAGCCAGAACGCTTTTCCGTCGTGTAGACAGCAACCGTTCGGTGTCTACAGC 285
 Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 Db 286 CGTAACGGCAATTTAAGCTGGATGAAACCGTAACCTGGTGAATATGCAAGTTTACAG 345
 Qy 129 LeuMetGlySerArglleSerAsnAsnProAsnileLysLysGluThrLeuGluProlle 148
 Db 346 CTGACGGGTTACCGGCAACCGGTACGCGCCGCACTATTACAGAGGGGGGAATTCGACC 405
 Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
 Db 406 AATATTTCCGATCCGATACCTCTGATGCA-----GCGAAACTACCAACCGCA 456
 Qy 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 Db 457 TCGATGCAGATCAACCTG-----AATTCAGTGTCCGCTTCTCTACT----- 498
 Qy 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProProlleSerThrSerAsn--- 207

Db	499	-----GTTACGCCATTCAGCGCCAGCAATGCG	525
Qy	208	-----TyrSerTyAlaGlnProMetArgValTyAspGlnGlnGlyAsnSerHisAsp	225
Db	526	GATAGCTATAACAAAAAGGTTGGTACTGTTTCGACATCGAGGTAATGCTCATGAC	585
Qy	226	lleThrValTyPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyLeu	245
Db	586	ATGAGCGTCTACTTT-----GTGAGACCGGGGATATACTGGCAG-----	627
Qy	246	ValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeu	265
Db	628	GTCTACACCCAGGATAGCAGTATCAACAGCACTTCGGAAGACA-----	672
Qy	286	LeuMetSerGlyThrMetThrPheSerSerSerSerGlnGlyLeuLysAsnMetThrAlaPhe	285
Db	673	-----GGCACAACCTCGAATTTAATGCTAATGCACTTA-----GTGGATGGTGGC	720
Qy	286	ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn	305
Db	721	ATGGCGAATAATATCGCAACCGGCGCAATTAACGGTGCAGAACCGGCC-----	768
Qy	306	GlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAsp	325
Db	769	-----ACGTTTAGTCTGAGCTTCTCACTCC-----	795
Qy	326	PheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIle	345
Db	796	-----ATCAGCAAAAT-----	807
Qy	346	GlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSer	365
Db	808	-----ACCGGGCGCTAACCAATATT	825
Qy	366	ThrAlaArgAsnGlySerSerSerThrArgArgTySerGlnAspGlyTyProGlnGly	385
Db	826	GTGGCAACC-----ACCGAAGCGGTACAAACCGGCG	858
Qy	386	AspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTySerAsnSer	405
Db	859	GATCTGGTCAATCAATCAATGATGACGTGCGTGTGGCAACTATTTCACCAAA	918
Qy	406	GlnValValAspPheTyAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg	425
Db	919	CAACCACTCTGGGCGAGATTGCTGGCGAACTTGCACCAACGAAGGTCTGGCA	978
Qy	426	ArgGluGlyAsnAsnHisTySerAlaThrLeuAspSerGlyTyProGluPheGlyLeu	445
Db	979	TCGGAAGCGCAACAGTCTGGTCTGGCAGCGCAATCTCTGGGTGGCGTGTGGGACA	1038
Qy	446	ProGlyThrSerAsnTyGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAsp	465
Db	1039	GCGGACGCGGAACTTGGCACTGACCAACCGTGGCGTGGAAAGCGTCCAACGTCGAT	1098
Qy	466	MetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLys	485
Db	1099	CTCAGTAAGAAGATGCTGATATGATGCTGGTGGCGAGCACTATCAGTCAACGCCGAG	1158
Qy	486	SerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys	501
Db	1159	ACCATCAAAACCCAGGACCATCTCTCAACAGCGTGGTTAACTTACGC	1206

RESULT 12
ABQ69245_07
Continuation (8 of 31) of ABQ69245 from base 700001 (Listeria innocua DNA sequence #694.
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245
WP Fragment Name Begin End
WP ABQ69245_00 1 110000
WP ABQ69245_01 100001 210000
WP ABQ69245_02 200001 310000
WP ABQ69245_03 300001 410000
WP ABQ69245_04 400001 510000

WP	ABQ69245_05	500001	610000
WP	ABQ69245_06	700001	710000
WP	ABQ69245_07	700001	810000
WP	ABQ69245_08	800001	910000
WP	ABQ69245_09	900001	1010000
WP	ABQ69245_10	1000001	1110000
WP	ABQ69245_11	1100001	1210000
WP	ABQ69245_12	1200001	1310000
WP	ABQ69245_13	1300001	1410000
WP	ABQ69245_14	1400001	1510000
WP	ABQ69245_15	1500001	1610000
WP	ABQ69245_16	1600001	1710000
WP	ABQ69245_17	1700001	1810000
WP	ABQ69245_18	1800001	1910000
WP	ABQ69245_19	1900001	2010000
WP	ABQ69245_20	2000001	2110000
WP	ABQ69245_21	2100001	2210000
WP	ABQ69245_22	2200001	2310000
WP	ABQ69245_23	2300001	2410000
WP	ABQ69245_24	2400001	2510000
WP	ABQ69245_25	2500001	2610000
WP	ABQ69245_26	2600001	2710000
WP	ABQ69245_27	2700001	2810000
WP	ABQ69245_28	2800001	2910000
WP	ABQ69245_29	2900001	3010000
WP	ABQ69245_30	3000001	3111208

Alignment Scores:
Pred. No.: 3e-27 Length: 110000
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-1 (1-502) x ABQ69245_07 (1-110000)			
Qy	1	MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly	20
Db	29803	ATGAATCAAACTATGATACAGCTATTTCGGATGATGATCGTTCACACAGCATTTATCA	29862
Qy	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyLysGlnGlnValVal	40
Db	29863	GTGCATCAATAATATTGCCAATGCGAACACACAGGATATAAAACAAACGCGTGGT	29922
Qy	41	PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer	56
Db	29923	TTCAATGATTACTTTTACCAAAACACAAATGGGATCTCTTCAGCGGACTTTATGCTGA	29982
Qy	57	GlnGlyProAsnGlnAlaGlyMetGlyValaGlnValGlySerValArgThrIlePheThr	76
Db	29983	ACAAACCAATGAGCTTCGGTTCGGTTCGAAATTCGGCAATTTTAAACGATTATACA	30042
Qy	77	GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe	96
Db	30043	GCAGGTTCTCCGACATCAACTGGCAGAAACAAAGATGCACACTGCAAGCGCGGCTTT	30102
Qy	97	PheGlnVal-----ThrLeuGluAspLysValHisTyThrArgAlaGlyAsnPheArg	114
Db	30103	TTCAATGCTGGCATAACGCTGGTGGGAATATCGTTTACACAGACGCGTGGCTTGA	30162
Qy	115	PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle	134
Db	30163	GTATCTGACATAACTATTAAACCAACCAACAGAAATACGTTATGGGATATGCAACA	30222
Qy	135	SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp	154
Db	30223	GATAAAATGCAACGTTTAAACGGAACCTTGCACCAATTCATATCCCATTAATAGC	30282
Qy	155	ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu	174
Db	30283	GCAATTCAGCGCAAGCAACAAATAGTGTAGCTTAAGCGGTAACATTCACCTTGTGG	30342

```
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
|||
Db 30343 GCGGAAAAA---GATACGATTCTTCGCGAG----- 30369
QY 195 TrpLysGlyAsnGlyThrProPheThrSerAsnTyrSerTyrAlaGlnProMet 214
|||
Db 30370 -----CTT 30372
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
|||
Db 30373 TCTGTATACGATATGCTGCGGAAAAAATAAACTTCAAGTCAATATGAAGCTGCGACA 30432
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
|||
Db 30433 CCAGATCGAGTGGCAATGTTCTTGAATATGAATCAAATG-----GAC 30480
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
|||
Db 30481 GGAAGCGTTAACTCTCCA-----GTAACAGCAACACTTAATTAC 30522
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
|||
Db 30523 AACCGCAAGCGCAACTTACAAACCTGATGCA----- 30555
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
|||
Db 30556 ---CTTAAACCATTCAAATCAATTCCTGTTTAAACGCG----- 30591
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
|||
Db 30592 -----AAACAAGTCAAT 30603
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
|||
Db 30604 ATG-----GGCTTAACCTAAGTGGCTTAAACA 30630
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
|||
Db 30631 AACTAC-----GGTACCAACCAA 30648
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
|||
Db 30649 GTATTCACCAACTCTCAGCGAAAGCGCTGCAACTGTAAAGACTACCGAGTTACC 30708
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
|||
Db 30709 GATTCGGTTATATTGACGTAGTACTCAGCGGTACAGTTATCCAGTTGCCCAACTT 30768
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
|||
Db 30769 GCGGTGCTACTTCTCCAAATGAAGCGCTTAGTCAAAATGGGAATGGCGAATATGTT 30828
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
|||
Db 30829 CCAGGATTATCTTCGAGATGCAAGTTACGCGGTGTGCGCAAAACGAGCTGGCGGA 30888
QY 454 LeuSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArgGluMetValAsnMet 473
|||
Db 30889 ATTAGCGGCTCTTCATTAGAAGGCTCAAAACGTAGATTTATCCCGGAATTCGTAACTTA 30948
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
|||
Db 30949 ATGACATACCAAGTGGTTTCAAGGCAATACAAAGTTATCCGTTGTCAGATGACGCTG 31008
QY 494 LeuGlnLysAlaLeuGluLeu 500
|||
Db 31009 ATCAACAAATTTGTAACCTG 31029

RESULT 13
AB067197.06
Continuation (7 of 12) of AB067197 from base 600001 (Listeria innocua contig DNA sequenc
WP Sequence split into 12 fragments LOCUS AB067197 Accession AB067197
WP Fragment Name Begin End
```

```
WP AB067197.00 1 110000
WP AB067197.01 100001 210000
WP AB067197.02 200001 310000
WP AB067197.03 300001 410000
WP AB067197.04 400001 510000
WP AB067197.05 500001 610000
WP AB067197.06 600001 710000
WP AB067197.07 700001 810000
WP AB067197.08 800001 910000
WP AB067197.09 900001 1010000
WP AB067197.10 1000001 1110000
WP AB067197.11 1100001 1163020

Alignment Scores:
Score: 3e-27 Length: 110000
Pred. No.: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-i (1-502) x AB067197.06 (1-110000)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
|||
Db 39538 ATGAATCAAACTATGATACAGCTATTTCTGGGATGATGCGTTCACACAGCATATCA 39597
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
|||
Db 39598 GTGACATCAAAATATATGCAATGCGAACACGACGAGGTATATAAACAAAGCGTCT 39657
QY 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
|||
Db 39658 TTCAATGATTTACTTTACCAAAACACAACTGGATCTGTTCCAGCGGCACTTTATGCTGGA 39717
QY 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
|||
Db 39718 ACAAACCAATGAGTTCGGTTCGGTTCGAAATTTGGGCAATTTTAAACCGATTATACA 39777
QY 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
|||
Db 39778 GCAGTTCTCCGACATCAACTGGCGAGAAACAAAGATGCACTGCAAGCGCGGCTTT 39837
QY 97 PheGlnVal-----ThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArg 114
|||
Db 39838 TTCATTCTCGCATACGCTGGTGGGAATATCGTTTACACAGCAGACGCTAGCTTTGCA 39897
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
|||
Db 39898 GTATCTGACATAACATATTATTAACACGCAACGCAACGAAATACTGTTATGGGATATGCAACA 39957
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
|||
Db 39958 GATAAAATGCGCAACGCTTTAAACGGAACCTTGCACCACTTCAATCCCATTAATAGC 40017
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
|||
Db 40018 GCAATTCAGCGGCAAGCAAAATAATGTTAGCGGTAAACATTCCTGATTTGG 40077
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
|||
Db 40078 GCGGAAAAA---GATACGATTTCTTCGAG----- 40104
QY 195 TrpLysGlyAsnGlyThrProPheIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
|||
Db 40105 -----CTT 40107
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla--- 233
|||
Db 40108 TCTGTATACGATATGCTGCGGAAAAAATAAACTTCAAGTCAATATGAAGCTGCGACA 40167
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
|||
```

Db 40168 CCAGATGGCAGTGGCAATGTTTCCTATGAAATGAAATTCAAATG-----GAC 40215
Qy 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 40216 GGAAGCGTTAACTCTCCA-----GTAACAGGAACACTTAATATAC 40257
Qy 274 SerSerAsnGlyGluLeuLysMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 40258 AACGGCAGGCGAACTTACAAACCTGATGCA----- 40290
Qy 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 40291 ---CITAAAAACATCAATCAATCCACTGTTACGGC----- 40326
Qy 314 PheValGlyAlaGlyLeuGlnProLeuThrLeuAspPheGlyLeuSerGlnGlnAsn 333
Db 40327 -----AAACAAGTCAAT 40338
Qy 334 MetTrpAlaGlyAlaProAlaSerAlaAlaAlaGlyThrAspLeuLysLeuPro 353
Db 40339 ATG-----GGTTAAACCTAAGTGGCTTACA 40365
Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 40366 AACTAC-----GGTACCAACCAA 40383
Qy 374 ThrArgArgTySerGlnAspGlyTyProGlnGlyAspLeuValAspValThrIleThr 393
Db 40384 GTATTCACCAACTCTCAGCGAAGCGGTGCAACTGTAAGACATACCGACTTACC 40443
Qy 394 SerGluGlyLeuLeuGlnGlyTySerAsnSerGlnValValAspPheTyAsnIle 413
Db 40444 GATTCGTGTTATATGACGTAGTTACTCAGACGGTACAGTTATCCAGTTCGCCCACTT 40503
Qy 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTySer 433
Db 40504 GCGGTGTCTACTTCTCCAATGAAGACGCTTAGTCMAAATGGGAATGGCGAATATGTT 40563
Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyGlyLys 453
Db 40564 CCAGGATATCTCTGAGATGCAGTTTACGGGTGCTGGCCAAACGGACCTGCGGA 40623
Qy 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 40624 ATTAGCGGCTCTTCATTAGAGCTCAACAGCTAGATTATTCGCGGAATTCGTTAACTTA 40683
Qy 474 IleIleIleGlaArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 40684 ATGACATACCAAGTGGTTCCAGGCAATACAAAGTTATCCGTTGCCAGTGCAGTG 40743
Qy 494 LeuGlnLysAlaLeuGluLeu 500
Db 40744 ATGAACAAATTCGAACTTG 40764

RESULT 14

ABA03041_07
Continuation (8 of 30) of ABA03041 from base 700001 (Listeria monocytogenes EGD-e genome
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041

Fragment Name	Begin	End
WP ABA03041_00	1	110000
WP ABA03041_01	100001	210000
WP ABA03041_02	200001	310000
WP ABA03041_03	300001	410000
WP ABA03041_04	400001	510000
WP ABA03041_05	500001	610000
WP ABA03041_06	600001	710000
WP ABA03041_07	700001	810000
WP ABA03041_08	800001	910000
WP ABA03041_09	900001	1010000
WP ABA03041_10	1000001	1110000
WP ABA03041_11	1100001	1210000
WP ABA03041_12	1200001	1310000
WP ABA03041_13	1300001	1410000

WP ABA03041_14 1400001 1510000
WP ABA03041_15 1500001 1610000
WP ABA03041_16 1600001 1710000
WP ABA03041_17 1700001 1810000
WP ABA03041_18 1800001 1910000
WP ABA03041_19 1900001 2010000
WP ABA03041_20 2000001 2110000
WP ABA03041_21 2100001 2210000
WP ABA03041_22 2200001 2310000
WP ABA03041_23 2300001 2410000
WP ABA03041_24 2400001 2510000
WP ABA03041_25 2500001 2610000
WP ABA03041_26 2600001 2710000
WP ABA03041_27 2700001 2810000
WP ABA03041_28 2800001 2910000
WP ABA03041_29 2900001 2944528

Alignment Scores:

Pred. No.: 3e-27 Length: 110000
Score: 484.50 Matches: 137
Percent Similarity: 42.21% Conservative: 77
Best Local Similarity: 27.02% Mismatches: 188
Query Match: 18.74% Indels: 105
DB: 6 Gaps: 11

US-10-009-823A-1 (1-502) x ABA03041_07 (1-110000)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 30516 ATGAATCAACTATGTAATACAGCTATTTCTGGATGATGATCGTTCACACAGCATTTATCA 30575
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyLysGlnGlnValVal 40
Db 30576 GTAACATCAATAATATTGCCAATGCCAACACACAGCATATAAAACAAACAGCGTGGTT 30635
Qy 41 PheGlnAspLeuPheSerGlnAsp-----LeuAlaIleGlySerThrGlySer 56
Db 30636 TTCAATGATTTACTTTTACCAAAACACAATGGGATCTGTCAGCGGACTTTATGCTGGA 30695
Qy 57 GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThr 76
Db 30696 ACAACCCCAATGAGTTTCGGTTCGGTTCGAAATTCGGGCGATTTTAACTGACTATACA 30755
Qy 77 GlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhe 96
Db 30756 GCAGGTTCGCCGACTTCACTGCGAGAAATAAAGATGCACACTACAAAGCCGCGGGCTTT 30815
Qy 97 PheGlnVal-----ThrLeuGluAspLysValHisTyThrArgAlaGlyAsnPheArg 114
Db 30816 TTCAATGCTGGTGTATACGCTGGTGGGAATATCGTTTATACAGTGCAGGTAGTTTGA 30875
Qy 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db 30876 GTATCTGACATAACTATTTAACACACGACAGCAAGAAATACGTTATGGGTACGCAACG 30935
Qy 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 30936 GACAAAATGTAACGTTTAAACGGTAATTTACACCCATCCAAATCCGCTAAATAGC 30995
Qy 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 30996 GCAATTCAGGGGAAGCAACAAAATAATGGTAGCCTAAGCGGTACATCCCACTGATTGG 31055
Qy 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyThrPheAlaLeuLeuGluSer 194
Db 31056 GCGCAAAAA---GATACGATTCTTCGAG----- 31082
Qy 195 TrpLysGlyAsnGlyThrProIleSerThrSerAsnTySerTyAlaGlnProMet 214
Db 31083 -----CTG 31085
Qy 215 ArgValTyAspGlnGlnGlyAsnSerHisAspIleThrValTyPheAspGlyAla--- 233

PF 23-JUN-1998; 98WO-US013041.
 PR 24-JUN-1997; 97US-0050667P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 FA Fraser CM;
 XX WPI; 1999-081273/07.
 XX New isolated Treponema pallidum nucleic acids - used to develop products
 PT for the detection, diagnosis, characterisation, prevention and therapy of
 PT T. pallidum infections, particularly syphilis.
 XX Claim 1; Page 491-497; 1150pp; English.
 XX AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection diagnosis,
 CC characterisation, prevention and therapy for T. pallidum infections,
 CC particularly syphilis. They can also be used for detecting diseases
 CC related to Borrelia infections in animals, and for the production of
 CC biosynthetic products such as enzymes
 XX Sequence 10461 BP; 2325 A; 3411 C; 2493 G; 2213 T; 0 U; 19 Other;
 SQ

Alignment Scores:
 Pred. No.: 7,67e-28 Length: 10461
 Score: 476.00 Matches: 153
 Percent Similarity: 42.34% Conservative: 79
 Best Local Similarity: 27.92% Mismatches: 185
 Query Match: 18.41% Indels: 134
 DB: 2 Gaps: 14

US-10-009-823A-1 (1-502) x AAX20553 (1-10461)
 QY 1 MetMetGlySerLeuPheLeuAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 2293 ATGATCGGTCACCTTTTTCAGGTGTCTGATGAGATCATCAACGGCGATGAT 2234
 QY 21 ThrValSerAsnAsnLeuAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 2233 GTCATTGGGAGAACGTCGCGAAGCGTTAACTACCTACCGGTTTAAAGCGTGGGTGTTAT 2174
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer-----ThrGly 55
 Db 2173 TTTCAGATCTTATTTCTGACCACTGAGTGGCGGTGCGCGTCCGCAATGCAAGATTGA 2114
 QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
 Db 2113 GGAGTGAATCCCAAGGAAGTGGGATTTGGCGTCTGATTGCAAGCATCATCTGTTTCA 2054
 QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLysGly 95
 Db 2053 ACCAAGTGCATGCAACAGCGGTATCAATACCGATGTGTCTATTCAGGGAGTGT 1994
 QY 96 PhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPhe 115
 Db 1993 TTTTGTGTGTAAGATGGGGAAGACGCTTTTTCACCGCGCAGT-GCCTTTGGGGT 1935
 QY 116 ThrGlnAspGlyPheLeuAsnAspProSer-----Gly 126
 Db 1934 GATAATCGCGGCACTCTGTAACCTCGCAATGATGATGCGGTTCAGGTTGGATGGCG 1875
 QY 127 PheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlnThrLeuGlu 146
 Db 1874 CAGGACGTGGCGGGGCGGTTTAAATTCCTCTGCACAGCAGCAGGATCTGTTATC 1815
 QY 147 ProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThr 166
 Db 1814 CCCATTGG-----GCAAAAGATAGATGCGCACAGACCCAGC 1780
 QY 167 AlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsn 186

Db 1779 ACTGTTCACTATGCTGTAAATTTTACACAAGCGTCTGCTGAGCTTGTGTCAGATGCGAAC 1720
 QY 187 ProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrProIleSerThrSer 206
 Db 1719 GAAGGGACGTGCTAAGTCCACGTG-----ACAACT 1687
 QY 207 AsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIle 226
 Db 1686 GACTTT-----CAAGTGTATGATAGCTTCGGCAGCAGCATACGTTG 1645
 QY 227 ThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVal 246
 Db 1644 CAGATTAACTTTTCGCTGTCGCGGAGC-----AACAACTAGTGGCAGCAGCATCTGC 1591
 QY 247 AlaMetAsnPro-----SerGluAsp 253
 Db 1590 GCAGTGGATCGGGGACAGAGGTAGATACGAAACGGTGTAGGGGTGGGACATCTGAC 1531
 QY 254 GlySerAla-----AlaSerGlyThr 260
 Db 1530 GGTGCGCAACACCTTTATTTGATAATTTTGGACACCTCGCTTCAGTGACT 1471
 QY 261 AspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 Db 1470 GACTGTCAGG----- 1459
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaThrGlnPro 300
 Db 1458 AACGTGACCGGT-----CCTACCGGACAGGTG----- 1432
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 Db 1431 -----CTCCTTGAAGCGTCTGATGTTGTCGTGGG----- 1399
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
 Db 1398 -----AATCCGACGATGACGGGAGGTT 1375
 QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 Db 1374 ACGCGCACGCTTTCAGCTCACTTGGGTGAAAT----- 1339
 QY 361 SerSerGlyAsnSerThrAlaArgAsn-----GlySerSerSerThr 374
 Db 1338 -----GGCACCGCGCGCAATACGATTACGCGATTGCTGAACGCGACTATACC 1291
 QY 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
 Db 1290 AAAGC-TACCGGACGAGGAGGTACCGATGGGATATTGGAAATTTTAAATAGATCA 1232
 QY 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
 Db 1231 AGCGTGTCTACTCTGTGTGTTTCAATGGGTGAGCCAGACATTTGCCAGCTCGCA 1172
 QY 415 LeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAla 434
 Db 1171 CTTGACGAGTTTCAATCAAGGTGCTTTGAGAAGCGACGAGAGACACCTACGTACAA 1112
 QY 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
 Db 1111 TCGAACACTCAGGGATAGCAACATTAGCACCTCGGGGTGATGGGAGGGAAGTTG 1052
 QY 455 SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle 474
 Db 1051 ATTGACGGGACACTTACGATGACGACAGTAGATTAAACCGATCAATTTACGGATATGATC 992
 QY 475 IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu 494
 Db 991 ATTACCCAAAGAGGTTTACGGCGGCGCAAGACGATTTCAGACATCAGACCATGTTG 932
 QY 495 GlnLysAlaLeuLeuLysArg 502
 Db 931 GATACGTTGTTGAGTTTGAAGCGC 908

RESULT 17

ADP03725 standard; DNA; 1239 BP.

ADP03725;

12-FEB-2004 (first entry)

Bacterial polynucleotide #4010.

Proteus mirabilis infection; bacterial infection; antibacterial;

immunostimulant; gene; ds.

Proteus mirabilis.

US6605709-B1.

12-AUG-2003.

05-APR-2000; 2000US-00543681.

09-APR-1999; 99US-0128706P.

(GENO-) GENOME THERAPEUTICS CORP.

Breton GL;

WPI; 2003-895291/82.

P-PSDB; ADP07897.

New Proteus mirabilis polypeptides and polynucleotides, useful as reagents for diagnosis of bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

Disclosure; SEQ ID NO 4010; 870pp; English.

The invention relates to new Proteus mirabilis polypeptides and polynucleotides. The invention also relates to antibodies against the polypeptides, methods for producing the polypeptides, a method of generating vaccines for immunising an individual against P. mirabilis, a method for evaluating a compound for the ability to bind a P. mirabilis polypeptide and a method for screening test compounds for anti-bacterial activity. The polypeptides and polynucleotides are useful as molecular targets for diagnosing, preventing and treating pathological conditions resulting from bacterial infection, as reagents for diagnosis of bacterial diseases, as components of antibacterial vaccines, as targets for antibacterial drugs or as bio-control agents for plants. This sequence represents a Proteus mirabilis polynucleotide of the invention.

Sequence 1239 BP; 398 A; 246 C; 262 G; 333 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.83e-29 Length: 1239
Score: 475.50 Matches: 136
Percent Similarity: 42.63% Conservative: 81
Best Local Similarity: 26.72% Mismatches: 165
Query Match: 18.39% Indels: 127
DB: 10 Gaps: 15

US-10-009-823A-1 (1-502) x ADF03725 (1-1239)

10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
40 AGTGGTTTAATGACGACGCGCTAATTAGACTACTATTGGTAATATATCTCAACTCT 99
30 AsnThrIleGlyTyrIysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49
100 GCAACCTACGTTTAAAGGGGCAACAGTCTCTTTTCCGATGTTTTTGC----- 150
50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69

151 -----GGCTCT-----GGTCAGGGCTTGGTGTAAAGATATCA 183
70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
184 GGCATTAGCCAAACTTTAAAGATGGTAGTACTACTACTAACCAGCCCAACCGATGTG 243
90 AlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLys-----ValHis 106
244 GCGATTTCTGGTGGTGGTTTCTTCCGTATT-----GAAGATAGTAACGGTGGCTTTTC 297
107 TyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGly 126
298 TATTTCAGTAACGGTGAATTTGGCAAGATAAAACGGCTATCTCACTAATATGCAAGGC 357
127 PheThrLeuMetGlySerArgIleSerAsn-----AsnProAsnIleLysLys 142
358 ATGGCTATTACAGGTATTCCAGTACAAAACGTAGACGGTAAATAATGTAGTACAAAAGGG 417
143 GluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAla 162
418 GCAACACCAACGCTCTATCATTT-----CCTACCGATATGATGAAT---GCA 462
163 LysThrSerThrAlaLeuAsnAlaValAlaAsnLeuGly-----AspSerThrAsp 179
463 AGTCAACCGATAAAATGGATATGACGGTTAACTTGAATTCAGCTGAAGAGGCTATCGAT 522
180 LysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGly 199
523 CAAACAACACATAAAATTTGATCCC-----AAAGATAAC--- 555
200 ThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGln 219
556 -----GATTCCTTATACTTAGTACTAAGTGACACACCTACGATAGC 597
220 GlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySer 239
598 TTAGGTAATGAACATAACCTGAATTTCTTT-----GTTAAAC 639
240 LysThrPheGlyTyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly 259
640 AAAGATACGANTGGAGTGTATTGCTCAAGACACACACAGGTGAGCCGACACAGAT 699
260 ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGly----- 277
700 CTG-----GGTAAATAGTTTATAGGATAACGGGTATTA 735
278 -----GluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThr 292
736 GATGAACACAGCACCTAAGCTGAAACCTTTACTACCGTTGCTTATAAGGTTCA----- 789
293 LysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAla 312
789 ----- 789
313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332
790 -----CAACCCATGATATGGAATGAACCTTT----- 816
333 AsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352
816 ----- 816
353 ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 372
817 -----AGTGGTAGTACGACGACAAAAGTGCCTGAATCA 849
373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392
850 AGTGTCTTAACCTCGGCAAAATGGTTTATCAACGAGGGGATTTACCAATTTCCGTATT 909
393 ThrSerGluGlyLysLeuGlnGlyTyrSerAsnSerGlnValValAspPheTyrAsn 412
910 GAGCCAGATGTTTCCATTATGCGACCTACTCAAAACCAACAAAGCCAAAGTGGTGTCA 969

QY 30 AsnThrIleGlyTyrIleGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
DB 90512 CGGACATAGGTTTAAATACGACCGGTGCTTTTCTGACATTTTGGC----- 90462
QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
DB 90461 -----GGTTCT-----CAGGTGGGACTGGGGTAAAGATTCG 90429
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
DB 90428 GGTGTTTAAATCAGAAATTTTAAAGATGTACGCCAACGACGACTAACCGAACACTTCATCTG 90369
QY 90 AlaIleGlyGlyGlyPheGlnValThrLeuGluAsp---LysValHisThr 108
DB 90368 GCGATCACCGCGGGGTTTTCGTATGCAAGCGGTGATGCGGGGATTTATTATTC 90309
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
DB 90308 CGTAACGGCCAATTCAGATGATGATGAGACCGCAATATTATCAACATGCAAGGCATGAAA 90249
QY 129 LeuMetGly-----SerArgIleSerAsnAsnProAsnIleLysLys---Glu 143
DB 90248 TTGACCGGTATCCCGCTCCGACGCTCAATGTGTGGCGCGCGAGATTTCAGAAAGGGCT 90189
QY 144 ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer-----Pro 161
DB 90188 ACCCGAGACCAATT-----ACATACCTCAAGTATATTGATG 90150
QY 162 AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
DB 90149 GCTAAAGCAACAACTAAGGTACGATGACGACCAACCTT---AATTCGATGATAAAAA 90093
QY 182 GlnSerGluAlaAsnProThrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
DB 90092 CGACAGAGGCT-----TTCTCG 90075
QY 202 ProIleSerThrSerAsnTyrSertyrAlaGlnProMetArgValTyrAspGlnGly 221
DB 90074 CCGGAAATAACGATCTTATACTATCTCCAGGTGTACCACCTTTCATAGCTTGGT 90015
QY 222 AsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThr 241
DB 90014 ACCCACATGCTATCCAGGTTTCTTT----- 89988
QY 242 PheGluTyrLeuValAlaMetAsnProSerGlu-----AspGlySerAla 256
DB 89987 -----GTGAAACGATGATAACGAATGCGAGGTTTCATGCTCAGGATAGTGGC 89940
QY 257 AlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsn 276
DB 89939 GTTGAGGCGCTAAAGCGCTTCGTTG-----GGGACGTTGAATTTGATGCGCAC 89889
QY 277 GlyGluLeu-----LysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
DB 89888 GGTAAACTAATGGTGACATAATACGTTACATTTAAT----- 89850
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
DB 89849 -----GTTCTCTCTCTGAAT----- 89835
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnAsn 333
DB 89834 -----GGTTCTAATGCGGTGATACACTATTGATTTAATGGCAGTAAACAGCAAAAG 89781
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
DB 89780 -----ATTGATACTGACGACATATCAGCACCC 89754
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
DB 89754 ----- 89754

QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
DB 89753 -----AAGCAAGACGGTTATTCGCCGAGGTGAATTCAGGGCTACCGTATTGAG 89706
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
DB 89705 CAAGATGCGGTGTCATCGGCACCTATACCAATGAGCAATCCCAATATTATTAGGTACATC 89646
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
DB 89645 GTTCTGACTAATCTTGTAAACCCAGAGGACTGCGAGCGAAGGTTGATAACGTTGGGTA 89586
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
DB 89585 GAGACGAGCGCTTCAGGTAACCCAGTGTGTCGATTTCGCGGCTTGGCGCAA 89526
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
DB 89525 TTGCTCAGTGTGTCGTGGAAGCATCCACGTCGATATGAGCCAGGAGCTGGTTAATATG 89466
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
DB 89465 ATCGTGGCGAGCGCAACTACCAATCCAGTGGCGCAACCACTAAACCCAGATCAGATC 89406
QY 494 LeuGlnLysAlaLeuGluLeuLys 501
DB 89405 CTACAAACGCTGGTCAGCATGCGC 89382

RESULT 21
ACF67367_12/c
Continuation (13 of 57) of ACF67367 from base 1200001 (Photorhabdus luminescens nucleotic
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367
Fragment Name Begin End
WP ACF67367_00 1 110000
WP ACF67367_01 100001 210000
WP ACF67367_02 200001 310000
WP ACF67367_03 300001 410000
WP ACF67367_04 400001 510000
WP ACF67367_05 500001 610000
WP ACF67367_06 600001 710000
WP ACF67367_07 700001 810000
WP ACF67367_08 800001 910000
WP ACF67367_09 900001 1010000
WP ACF67367_10 1000001 1110000
WP ACF67367_11 1100001 1210000
WP ACF67367_12 1200001 1310000
WP ACF67367_13 1300001 1410000
WP ACF67367_14 1400001 1510000
WP ACF67367_15 1500001 1610000
WP ACF67367_16 1600001 1710000
WP ACF67367_17 1700001 1810000
WP ACF67367_18 1800001 1910000
WP ACF67367_19 1900001 2010000
WP ACF67367_20 2000001 2110000
WP ACF67367_21 2100001 2210000
WP ACF67367_22 2200001 2310000
WP ACF67367_23 2300001 2410000
WP ACF67367_24 2400001 2510000
WP ACF67367_25 2500001 2610000
WP ACF67367_26 2600001 2710000
WP ACF67367_27 2700001 2810000
WP ACF67367_28 2800001 2910000
WP ACF67367_29 2900001 3010000
WP ACF67367_30 3000001 3110000
WP ACF67367_31 3100001 3210000
WP ACF67367_32 3200001 3310000
WP ACF67367_33 3300001 3410000
WP ACF67367_34 3400001 3510000
WP ACF67367_35 3500001 3610000
WP ACF67367_36 3600001 3710000
WP ACF67367_37 3700001 3810000
WP ACF67367_38 3800001 3910000
WP ACF67367_39 3900001 4010000

WP ACF67367_40 4000001 4110000
 WP ACF67367_41 4100001 4210000
 WP ACF67367_42 4200001 4310000
 WP ACF67367_43 4300001 4410000
 WP ACF67367_44 4400001 4510000
 WP ACF67367_45 4500001 4610000
 WP ACF67367_46 4600001 4710000
 WP ACF67367_47 4700001 4810000
 WP ACF67367_48 4800001 4910000
 WP ACF67367_49 4900001 5010000
 WP ACF67367_50 5000001 5110000
 WP ACF67367_51 5100001 5210000
 WP ACF67367_52 5200001 5310000
 WP ACF67367_53 5300001 5410000
 WP ACF67367_54 5400001 5510000
 WP ACF67367_55 5500001 5610000
 WP ACF67367_56 5600001 5648894

Alignment Scores:
 Pred. No.: 2,41e-24 Length: 110000
 Score: 448.50 Matches: 141
 Percent Similarity: 40.75% Conservative: 66
 Best Local Similarity: 27.76% Mismatches: 174
 Query Match: 17.34% Indels: 127
 DB: 10 Gaps: 17

US-10-009-823a-1 (1-502) x ACF67367_12 (1-110000)

QY 10 ThrGlyMetLysThrHisSerAsnLeuGlyThrValSerAsnLeuAlaAsnAla 29
 DB 27341 AGTGGTTGAATGCCCGGAGGTAATTGGATGTTATGTAATAATATCGCTAACGCG 27282
 QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
 DB 27281 CGGACATATGTTTAAATCCAGCAGCGTCTTTCTGCATTTTGGC----- 27231
 QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
 DB 27230 -----GGTTCT-----CAGTGGGACTGGGGTAAAGTATCG 27198
 QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
 DB 27197 GGTCTTAATCAGATTTTAAAGTGTACCCACGACGACTAACCGAACACTTGTATCG 27138
 QY 90 AlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThr 108
 DB 27137 GCGATCAGCAGGCGGCTTTTTCGATGCAAGCGGCGTATCGGCGGATTTATTATTC 27078
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 DB 27077 CGTAACGCCCAATTCAGATGATGATGACCGCAATATATCAACATGCAAGCATGAAA 27018
 QY 129 LeuMetGly-----SerArgIleSerAsnAsnProAsnIleLysLys---Glu 143
 DB 27017 TTGACCGCTATCCGCTGCGCAGCGTCAATGTGTGGCGCGCGAGATTCAGAAAGGGCT 26958
 QY 144 ThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSer-----Pro 161
 DB 26957 AACCCAGAGCAAT-----ACATACCTCAAGGTATATTGATG 26919
 QY 162 AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThr 181
 DB 26918 GCTAAACCAACAACTAAGGTACGATGACAGCCCACTT---AATTCGATGATATAAAAA 26862
 QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
 DB 26861 CCGCAGAGGCT-----TTCTCG 26844
 QY 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGly 221
 DB 26843 CCGGAATAACGATCTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 26784
 QY 222 AsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThr 241

DB 26783 AACCATGCTATCCAGTTTCTTT----- 26757
 QY 242 PheGlyTyrLeuValAlaMetAsnProSerGlu-----AspGlySerAla 256
 DB 26756 -----GTGAAAACCAATGATAACGAATGCGAGTTTCATGCTCAGGATGTGCG 26709
 QY 257 AlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsn 276
 DB 26708 GTTGAGGCGCTAAAGCGCTTCGTTG-----GGGACGTTGAATTTGATGGCAAC 26658
 QY 277 GlyGluLeu-----LysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 DB 26657 GGTAACTAATTTGTGACAAATATCAGTTTACATTTAAT----- 26619
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 DB 26618 -----GTTCTCTCTCTGAAT----- 26604
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 DB 26603 -----GGTTCTAATGCGCTGATATCATTGTTTAAATGCGCAGTAACACCAAAAG 26550
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 DB 26549 -----ATTGATCTGACAGCATATCAGCACCC 26523
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
 DB 26523 ----- 26523
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 DB 26522 -----AAGCAAGACGGTTATGCGCAGGTGAATTCAGGGCTACCGTATTGAG 26475
 QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 DB 26474 CAAGATGCCGTGTCATCGGCACCTATACCAATGAGCAATCCCATTTATTAGTTCAGATC 26415
 QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGlyAsnAsnHisTyrSer 433
 DB 26414 GTTCTGACTAATTGCTAACCCAGAGGAGTGCAGCAGGAGGTGATAACGTTGGGTA 26355
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 DB 26354 GAGACGAGCGCTTCAGGTAAACCCAGTGTGGCATTCGCGGATTCGCGGTTTGGCAAA 26295
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 DB 26294 TTGCTCAGTGTTCGCTGGAGCATCCACGTCGATATGAGCCAGGAGCTGTTAATATG 26235
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
 DB 26234 ATCGTGGCGCAGCGCAACTACCAATCCAGTGCAGCAACCACTAAACCCAGATCAGATC 26175
 QY 494 LeuGlnLysAlaLeuGluLeuLys 501
 DB 26174 CTACAAACGCTGGTCAGCATGCGC 26151

RESULT 22

ABA92787_3
 Continuation (4 of 7) of ABA92787 from base 300001 (Buchnera sp. genomic DNA SEQ ID NO:1.
 WP Sequence split into 7 fragments LOCUS ABA92787 Accession ABA92787

WP	Fragment Name	Begin	End
WP ABA92787_0	1	110000	
WP ABA92787_1	100001	210000	
WP ABA92787_2	200001	310000	
WP ABA92787_3	300001	410000	
WP ABA92787_4	400001	510000	
WP ABA92787_5	500001	610000	
WP ABA92787_6	600001	640681	

Alignment Scores:

Pred. No.:	1.33e-21	Length:	110000
Score:	414.50	Matches:	132
Percent Similarity:	39.96%	Conservative:	73
Best Local Similarity:	25.73%	Mismatches:	187
Query Match:	16.03%	Indels:	121
DB:	6	Gaps:	13

US-10-009-823A-1 (1-502) x ABA92787_3 (1-110000)

QY	4	SerLeupheileGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer	23
DB	72153	TCAAATATGATAGTAAAGTGGCTTACTAGCAAAATAATGATTACATGGAAATTTATATCC	72212
QY	24	AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAsp	43
DB	72213	ATAAATATGTTAAGCATCACTATAGGATATAATCTCGTAAACCTCTTTTGTAT	72272
QY	44	LeupheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly	63
DB	72273	ATGTTTCTCATTCATTTTATTCAATATCTACTAAT	72311
QY	64	MetGlyAlaGlnValGlySerValAsgThrIlePheThrGlnGlyAlaPheGluProGly	83
DB	72312	TACGGAGTGGGTATTTCAAGTATATACAAACTTTTAATAATGGCATGTAGTTGAACT	72371
QY	84	AsnSerValThrAspLeuAlaIleGlyLysGlyPhePheGln---ValThrLeuGlu	102
DB	72372	GGACGAGATTTGGATTTAGGAATTATAAAGACGGCTTTTTCGCTCTTGACACAGTCAA	72431
QY	103	AspLysValHisThrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn	122
DB	72432	GGTCATGCTTATATACAGAGATGGGCAATTTCTTCTCGATAAAGATCAAAATATATC	72491
QY	123	AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn	136
DB	72492	AATATTCAAGGTATGATCTAACTGGACTTAATACATCTTGTCAAAAAGTGAATTTAAT	72551
QY	137	AsnProAsnIleLysGlyThrLeuGluProIleGlnLeuAspPheAsnAspProThr	156
DB	72552	AAT-----AGATCCAACTTAGAACCTTATTAATTTA-----AAAAATCTAAT	72593
QY	157	ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly---	175
DB	72594	ATTTTAAAAACAACCTACTTCTGAAATCATGCTTAAAGCGTGTGTTGATCGTAACTAT	72653
QY	176	-----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe	189
DB	72654	GAATCAAAAAGCAGTGTGGTAAATTTCTGACACAAACTATCTAAACGAGAAGATTACATG	72713
QY	190	AlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSer	209
DB	72714	ACTTATATAGC-----	72725
QY	210	TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr	229
DB	72726	-----ATATATAATAAGAGGAGGAAAAAAGAGATATATTACTGTTTCT	72767
QY	230	PheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsn	249
DB	72768	TTTAAAT-----AAAAAGGAAACAATAATATGACAGTAAAT-----GTGGAATCAAT	72815
QY	250	ProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly	269
DB	72816	GATTCGTATGATAAGAGACTATAAAAAAT-----AGTTTC	72851
QY	270	ThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly	289
DB	72852	GATTTAACGTTTAAATGATGATGCGAATTAACCTCTGATAATGTTTAAATATTACATCT	72911
QY	290	SerAlaThrLysAspLeuAsnAlaTTPGlnProAlaProLeuValAsnGlyLeuProGln	309
DB	72912	AAAGATTTCTAAAG-----	72926

310 PheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLys 329
72927 -----TATGAAATATCATCTTTTAAATTTA----- 72950
330 SerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIle 349
72950 ----- 72950
350 GlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsn 369
72951 -----ACAGGTACTATAGAACATCAAT 72974
370 GlySerSerSerThrArgGlySerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389
72975 TCTGATGTTTCTTGGGAAGAACATCTCAAAAACGATACCTCAAGGTAATTTAAAAACA 73034
390 ValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAsp 409
73035 TTTGATATTGTTACTAATGGTGAATTTATGGAACATATTGCAATCAAAAACACAAACA 73094
410 PheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsn 429
73095 ATAGTCAAAATATTATTATCAAAATTTATCAATCCAGAAAAAATTACAACTGAAAGTGGT 73154
430 AsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu---ProGlyThr 448
73155 AATTATGGTCTGCTACTGCGAATCAGGTGAGCAAAACAGCAATGAAAGCGGTATT 73214
449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
73215 CAAGAATCAGGAGTGTAAAGCAATAAAACGCTAGCAAGTATCAAAATGTTGATTGCAATAA 73274
469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
73275 GAATTAATCAATATGATTATAGCAACACGTAATTTATCAATCTAACGCTCAATCTTTTAAA 73334
489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
73335 ACAGAGATAAATAATTAATACATTAAATTTACAG 73373

RESULT 23
AAS88896
ID AAS88896 standard; cDNA; 1912 BP.
XX AAS88896;
AC AAS88896;
XX 13-FEB-2002 (first entry)
DT DNA encoding novel human diagnostic protein #24700.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABC24709.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

QY 369 AsnGlySer-----SerSerThrArgArgTyrSerGlnAsp 380
 Db 94714 GAAGAGATGATTTTACAAATTTCTTACCTGATGACACTTATGATATATAGAT 94773
 QY 381 Gly----- 381
 Db 94774 GGGTCATTAAATCGATTTCTAATCGAGAGCTTGTACAAAGGATACAAAGTATTG 94833
 QY 382 -----TyrProGlnGlyAspLeuValAsp---ValThrIleThrSerGluGly 396
 Db 94834 CCTAATATATCTTCCAGAGAAATATATCCAAATCAATCAATATCTGAAGAGGA 94893
 QY 397 LysLeuGlnGlyLysTyrSerAsnSer---GlnValValAspPheTyrAsnIleProLeu 415
 Db 94894 ATAGTATCGGTAAATTTGATACCAAGCAACCAATAGAGCTTGGCAATTTGAATA 94953
 QY 416 AlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThr 435
 Db 94954 TCAAGATTATCAATCTCGAGGACTAAGTCCCATTCGAAGCAATTTATTTAAAGAAACA 95013
 QY 436 LeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSer 455
 Db 95014 GCTGATCAGGCCCAAGAAATAGCAGGAATACCAAGAGTGAAGCATGGGAAGACTAAGG 95073
 QY 456 ValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIle 475
 Db 95074 CAAGGCATCTTGAATGTCAATGTATCTATTGCTGAAGAAATGGTAACAATGATAGTA 95133
 QY 476 IleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGln 495
 Db 95134 GCTCAAGGGCTTATGAATAAATCAAAAGCTATTCAAACTTCTGACATATGTTAGGA 95193
 QY 496 LysAlaLeuGluLeuLysArg 502
 Db 95194 ATTGCAAACTAATTAAGG 95214
 RESULT 25
 ABD04540
 ID ABD04540 standard; DNA; 810 BP.
 XX
 AC ABD04540;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Pseudomonas aeruginosa polynucleotide #3144.
 XX
 DE Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
 KW antibacterial.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN US5551795-B1.
 XX
 PD 22-APR-2003.
 XX
 PF 18-FEB-1999; 99US-00252991.
 XX
 PR 18-FEB-1998; 98US-0074788P.
 PR 27-JUL-1998; 98US-0094190P.
 XX
 FA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
 XX
 DR WPI: 2003-615309/58.
 DR P-PSDB; ABO70969.
 XX
 DR Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of
 PT pathological conditions resulting from bacterial infection.
 XX
 PS Disclosure; SEQ ID NO 3144; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html
 XX
 SQ Sequence 810 BP; 166 A; 299 C; 226 G; 119 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.35e-15 Length: 810
 Score: 307.00 Matches: 108
 Percent Similarity: 30.74% Conservative: 46
 Best Local Similarity: 21.56% Mismatches: 93
 Query Match: 11.87% Indels: 254
 DB: 11 Gaps: 11
 US-10-009-823A-1 (1-502) x ABD04540 (1-810)
 QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 25 ATGTATCGGCACCTGGTCTGCGCAGCAACCGTCTGTCCGCCAGGACATGAACCTGACC 84
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 85 ACCATTTCCAAACAACCTGGCCCAAGCTATCCACCAACCGGCTTCAAGCGCGCGGAG 144
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 Db 145 TTCACGACCTTGTGTACCATCCGGCGCCAGCGCGGCGGCGGCGGCGGCGGCGGCGG 204
 QY 61 GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla 79
 Db 205 GAGCTGCTTCGGGCGCTGCAACTGGTACCGGTGTGGCGCTGTGGCGCACCCAGAGATC 264
 QY 80 PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLys 94
 Db 265 TTCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 324
 QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
 Db 325 GGCCTTCTCCAGGCTCTGCTCCCGGACGGCACCGCTGTCTACACCCGCGGCGGCGGCTTC 384
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 Db 385 CACCTGAACCTCCGACGGGCGGAGATCGTCACTCCACCGGCTTCGCCCTG----- 432
 QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 Db 432----- 432
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 Db 433 GAGCCAGCGGATC----- 444
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193
 Db 444----- 444
 QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 Db 444----- 444

QY	214	MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla	233
Db	444	-----	444
QY	234	ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp	253
Db	445	-----GTGGTGCCCAACGAG-----	459
QY	254	GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe	273
Db	460	-----ACCCAGACCTTC	471
QY	274	SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys	293
Db	472	ACC-----	474
QY	294	AspLeuAsnAlaTrpGlnProAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro	313
Db	474	-----	474
QY	314	PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn	333
Db	475	-----GTCGCGC-----	480
QY	334	MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro	353
Db	480	-----	480
QY	354	SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer	373
Db	480	-----	480
QY	374	ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr	393
Db	481	-----CAGGACGCG-----ACCGTCTCGGTGAC	504
QY	394	SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle	413
Db	505	ACCACCGCGCAACGCCAG-----CCGCGGTGATCGGC-----AACATC	543
QY	414	ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer	433
Db	544	CAGACCGCGACTTCATCAACCGCGCGCGCTCGAGCCCATCGCAACACCTGTCCTG	603
QY	434	AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys	453
Db	604	GAAACCGGCTCCAGCGCGCGCGCGAGTGGTACGCGGGTCTCAACGCGCTCGCACG	563
QY	454	LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet	473
Db	664	GTTGCCCAGAACACCCCTGGGAAACTCCACGTCAACGTGGTCGAGGAACCTGTTAACTG	723
QY	474	IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet	493
Db	724	ATCACCCACCGCGCGCTACGAGATGAATCCAGTCACTCCACCGCGCGACGATG	783
QY	494	Leu 494	
Db	784	TTG 786	
RESULT 26			
ABD04219/c			
ID	ABD04219 standard; DNA; 1545 BP.		
XX	AC ABD04219;		
XX	29-JUL-2004 (first entry)		
XX	Pseudomonas aeruginosa polynucleotide #2823.		
XX	Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;		
KW	antibacterial.		

XX	OS	Pseudomonas aeruginosa.
XX	PN	US6551795-B1.
XX	PD	22-APR-2003.
XX	PF	18-FEB-1999; 99US-00252991.
XX	PR	18-FEB-1998; 98US-0074788P.
XX	PR	27-JUL-1998; 98US-0094190P.
XX	PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	PI	Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX	DR	WPI; 2003-615309/58.
XX	DR	P-PSDB; ABO70648.
XX	PT	Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, of
XX	PT	useful as molecular targets for diagnostics, prophylaxis and treatment of
XX	PT	pathological conditions resulting from bacterial infection.
XX	FS	Disclosure; SEQ ID NO 2823; 455pp; English.
XX	CC	The invention relates to Pseudomonas aeruginosa polypeptides and the
XX	CC	polynucleotides encoding them. The sequences are useful in diagnosis and
XX	CC	therapy of pathological conditions, as molecular targets for diagnostics,
XX	CC	prophylaxis and treatment of pathological conditions resulting from a
XX	CC	bacterial infection, for evaluating a compound, such as a polypeptide, of
XX	CC	for the ability to bind a P. aeruginosa nucleic acid, as components of
XX	CC	effective antibacterial targets, as targets for antibacterial drugs,
XX	CC	including anti-P. aeruginosa drugs, as templates for recombinant
XX	CC	production of P. aeruginosa-derived peptides or polypeptides, as target
XX	CC	components for diagnosis and/or treatment of P. aeruginosa-caused
XX	CC	infection, and in detection of P. aeruginosa sequences or other sequences
XX	CC	of Pseudomonas species using biochip technology. Sequences ABD01397-
XX	CC	ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
XX	CC	The sequence data for this patent did not form part of the printed
XX	CC	specification but was obtained in electronic format from USPTO at
XX	CC	seqdata.uspto.gov/sequence.html
XX	SQ	Sequence 1545 BP; 226 A; 447 C; 568 G; 304 T; 0 U; 0 Other;

Alignment Scores:			
Pred. No.:	7,66e-15	Length:	1545
Score:	302.00	Matches:	107
Percent Similarity:	30.60%	Conservative:	46
Best Local Similarity:	21.40%	Mismatches:	93
Query Match:	11.68%	Indels:	254
DB:	11	Gaps:	11
US-10-009-823A-1 (1-502) x ABD04219 (1-1545)			
QY	2	MetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThr	21
Db	1545	TTATCGCAGCTGTGGTCAGCAAGACCGGTGTGTCGCCCGGAGCATGACCTGACAC	1486
QY	22	ValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPhe	41
Db	1485	ATTTCCAAACACCTGGCCCAACGTATCCACACCGGCTTCAAGCGCGCGGAGTTC	1426
QY	42	GlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGln	61
Db	1425	CAGGACCTGCTGTACCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1366
QY	62	AlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAlaPhe	80
Db	1365	CTGCCTTCGGGCGCTCAACTGGGTACCGGTGTGCGGTGCGGCGGCGGCGGCGGCGG	1306
QY	81	GluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLysGly	95
Db	1305	ACCCGGGCGAGCTGCAGACCAACCGGAGCGCGCTGGACATGGCGGTCAACGGCGCGG	1246

QY	96	PhePheGlnValThrLeuGluAsp	--LyeValHisTyrThrArgAlaGlyAsnPheArg	114
DB	1245	TTCTTCCAGGTCTGTGTCGGCAGCGCACCGTGTCTCTACCCGCGCAGCGACTTCCAC	1186	
QY	115	PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle	134	
DB	1185	CTGAACCTCCGACGGCAGATCGTCACCTCCACCGGCTTCGCCCTG	1141	
QY	135	SerAsnAsnProAsnIleLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp	154	
DB	1140	---	---	1138
QY	155	ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu	174	
DB	1137	CCAGCGATC	---	1129
QY	175	GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluGluSer	194	
DB	1129	---	---	1129
QY	195	TpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet	214	
DB	1129	---	---	1129
QY	215	ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro	234	
DB	1129	---	---	1129
QY	235	SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly	254	
DB	1128	---	---	1114
QY	255	SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer	274	
DB	1113	---	---	1099
QY	275	SerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp	294	
DB	1099	---	---	1099
QY	295	LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe	314	
DB	1099	---	---	1099
QY	315	ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet	334	
DB	1098	GTCGGC	---	1093
QY	335	TrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSer	354	
DB	1093	---	---	1093
QY	355	MetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThr	374	
DB	1093	---	---	1093
QY	375	ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer	394	
DB	1092	---	---	1066
QY	395	GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro	414	
DB	1065	ACCGGCAACGCCAG	---	1027
QY	415	LeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAla	434	
DB	1026	ACCGCGGACTTCATCAACCCCGCGCGCTCGAGCCATCGGCACACCTGTTCTCTGGAA	967	
QY	435	ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu	454	
DB	966	ACCGGCTCCAGCGCGCGCGAGGTCTGCTACGCGGGTCTCAACGCCCTCGCAGCGTT	907	

QY	455	SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle	474
	---		----
Db	906	GCCACGAAACCCCTGGAAAACTCCACGCTCGAGTGGTGGTGAACATGATC	847
	---		----
QY	475	IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu	494
	---		----
Db	846	ACCACCCAGCGCGCTACGAGATGAATCCAGGTGATCTCCACCGCCGCCAGATGTTG	787
	---		----
RESULT 27			
ID	ACF68481	ACF68481 standard; DNA; 783 BP.	
XX	AC	ACF68481;	
XX	XX	20-NOV-2003 (first entry)	
XX	XX	Photorhabdus luminescens nucleotide sequence #6948.	
DE	XX	Antibacterial; fungicide; insecticide; polymorphism; generic analysis;	
KW	KW	detection; food; gene expression; plant; animal; microorganism; toxin;	
KW	KW	antibiotic; biopesticide; virulence factor; disease model; plague;	
KW	KW	whooping cough; gene; ds.	
XX	XX	Photorhabdus luminescens.	
OS	XX	WO200294867-A2.	
PN	XX	28-NOV-2002.	
PD	XX	07-FEB-2002; 2002WO-IB003040.	
PF	XX	07-FEB-2001; 2001FR-00001659.	
PR	XX	(INSP) INST PASTEUR.	
XX	PA	(CNRS) CNRS CENT NAT RECH SCI.	
XX	PA	Duchaud E, Tacoutit S, Glaser P, Frangeul L, Kunst F, Danchin A;	
PI	PI	Buchrieser C;	
XX	XX	WPI; 2003-148459/14.	
DR	XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,	
XX	XX	useful e.g. as therapeutic antimicrobials and agricultural pesticides.	
PT	PT	Claim 2. SEQ ID NO 6948; 1205pp; French.	
PS	PS		

XX The invention relates to the isolation of genes and their encoded
CC proteins from *Photobacterium luminescens*. The isolated sequences are
CC sources of probes and primers for detecting the genome of *P. luminescens*
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than *P. luminescens* and are able to alter
CC response or sensitivity to toxins and antibiotics produced by *P.*
CC *luminescens*. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which *P.*
CC *luminescens* is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated *P. luminescens* genes
XX
SQ Sequence 793 BP; 235 A; 163 C; 178 G; 207 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.2e-14 Length: 793
Score: 295.00 Matches: 104

US-10-009-823A-1 (1-502) x ACF68481 (1-783)

Claim 1; Page 261-279; 1150pp; English.

RESULT 28
AAX20515/c
ID AAX20515 standard; DNA; 32768 BP.

05-MAY-1999 (first entry)

xx
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;

XX
ТРЕПОШКА ПАТРИЦА.

FD
XX
30-DEC-1998.

PR 24-JUN-1997; 9705-0030007F.
XX

XX
FLASER CM,

PT T. pallidum infections, particularly syphilis.

Claim 1; Page 261-279; 1150pp; English.

CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection, diagnosis,
 CC characterisation, prevention and therapy for T. pallidum infections,
 CC particularly syphilis. They can also be used for detecting diseases
 CC related to Borrelia infections in animals, and for the production of
 CC biosynthetic products such as enzymes

XX Sequence 32768 BP; 8253 A; 9783 C; 7257 G; 7433 T; 0 U; 42 Other;

Alignment Scores:

Pred. No.: 2,7e-12 Length: 32768
 Score: 291.00 Matches: 145
 Percent Similarity: 34.37% Conservative: 77
 Best Local Similarity: 22.45% Mismatches: 196
 Query Match: 11.25% Indels: 229
 DB: 2 Gaps: 22

US-10-009-823A-1 (1-502) x AAX20515 (1-32768)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 30751 ATGATCCAGGGTGGTATATCCGAGCCAGTGGCATGAGCGGCGAGCGCGCGTGGAT 30692
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 30691 GCTATTGGCGAAACCTAGCGAATGTAGACACAAAGCTACAGCGAGATGTAGCAGTT 30632
 QY 41 -----PheGlnAspLeuPheSerGlnAspLeu----- 49
 DB 30631 CACAGAGTTTTCAGAGCTCTATTGGCTGTAAACAGCATGGTGTGTGAAGAAC 30572
 QY 50 AlaIleGlySerThrGly-----SerGlnGlyProAsnGlnIleGlyMetGlyAlaGlnVal 68
 DB 30571 CCTTTCGGTCTCGGAGTTCTCTATAGTTGGAAGTTAGGCTCGGTGTAGGGTG 30512
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
 DB 30511 AATGAAGTTCTCAGTAATTTGAGAGGGTTCGCTTAAACAAAGCGAAGATCCTCGGAT 30452
 QY 89 LeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
 DB 30451 ATTGGCTTGGGGATCGGTTTTTTTGTGATTAGGACTCCACAGGGGAGAAGATACACC 30392
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 DB 30391 CGGAATGGGAATTTCTAGTCGGTGTAGAGGTTACCTTATGACTAAATGGGTATCCA 30332
 QY 129 LeuMetGlySerArg----- 137
 DB 30331 GTACTGGAGAAATGGTCCACTTTTTCGAGGAGGATACATACATCAATCAAAAC 30272
 QY 138 ProAsnIleLysLysGluThrLeuGluProIleGlnLeuAsp-----PheAsnAsp--- 154
 DB 30271 GGGGAGATTATGTACGTCCCATAGATAGGCGACAGCTTGACGGTTTTTTTTTTAGATCGG 30212
 QY 154 ----- 154
 DB 30211 CTGAAATTTGACGTTTGAATAATGTTGCTTACCTGCAAAAGAGGGTGTGATACGTAT 30152
 QY 155 -----ProThrValAlaLysSerPro----- 161
 DB 30151 ATGCAAACTCCTGTTTCAGGGGCCCTTATTGCTCGGAAGTCTCTGAACGTCGGGTGCA 30092
 QY 162 -----AlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly 175
 DB 30091 GTGCAGGGTTTTTGTAGCGCTCGAATGTAAACGTGGTCAATGAAATGGTTCTGATGATT 30032
 QY 176 AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAla-LeuLeuGluSerTr 195
 DB 30031 GAGGTGAATCGCGGTATGAGGCGAACCACAAATACTATTCAAGCAGAAGATGGAATGATG 29972
 QY 195 PLYSGlyAsnGlyThr----- 200

DB 29971 GGCAGATTGTGGAACGAAAGTGGTGGTGCATGCAAAAGTAGTTGTCCGCATGTTTTTATAGAATGT 29912
 QY 201 -----ProProIleSerThrSerAsn-----TyrSe 209
 DB 29911 AAAGAAGATATAGACGACATAGCGTGTAGCGGAGTGTAGAGGGAGGTGTGAATGCTACGA 29852
 QY 209 rTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTy 229
 DB 29851 ATTTCTGGAC-----CGTGTCTA 29834
 QY 229 rPheAspGlyAlaProSer-----SerThrGlySerLysThrPheGly 243
 DB 29833 CGGGGATGAACCGCCAGCAGCAACATCGATACGATGCGGAAATAATCTACGGAATGTGA 29774
 QY 243 uTyrLeu-----ValAlaMetAsnProSe 251
 DB 29773 ATACGTCTGGTTTTTAAAGACAGCGCTGCGGAGTTCGAGGACCTGTGTATCAGACCATAC 29714
 QY 251 rGluAspGlySerAlaAlaSer-----GlyThrAspSerAl 263
 DB 29713 GCACCGCGGTACGCTGCAACTGAGGACACACTTACCCCTGTGGGTACGGATGGGC 29654
 QY 263 a-GlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn--- 281
 DB 29653 ACGGGGTGAACCTCGCTCAACGCGAGCKTCTTCGACACAGGTTCCGTCAGCACACGG 29594
 QY 282 -----MetThrAlaPheThrP 287
 DB 29593 GTGTGAGCGGAGATCGCTATTGCGGTGAGGGTTTTTTCGCGTACTCAATATGACG 29534
 QY 287 roThrGlySerAlaThrLysasp-----LeuAsnAlaTrpGlnProAlaProL 303
 DB 29533 GAACCTATGCTTATACCCGCGAGCGGTTCCGTTCAAGTGGATGCGGAGCGCAATGGTGA 29474
 QY 303 euValAsnGlyLeu-----ProGln-----PheSerAlaAsnPheValGlyAlaG 318
 DB 29473 CCTCAACCGGTTGCTGCTCTCGTTCCAGAGATTATTTTCCGAAACTACGTAGAACACA 29414
 QY 318 lYleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyA 338
 DB 29413 GTATC----- 29409
 QY 338 laProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProI 358
 DB 29408 -----CCTATTAGCAAGAT----- 29394
 QY 358 leGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrS 378
 DB 29393 -----GGGCGCGTTACGTCGCGAGTCGGGAA----- 29367
 QY 378 erGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysL 398
 DB 29366 -----CAGGAGATCCGTTGAGTA-----GGACAGC 29339
 QY 398 euGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgp 418
 DB 29338 TTGAG-----CTGTA-TNCT 29325
 QY 418 heThrSerGluAspGlyLeuArgGlyGlyAsnAsnHisTyrSerAlaThrLeuAsps 438
 DB 29324 TCGCAACAACTCGGGGTTTCAGCAGAGGGGGGATGTGTTTACCAACACACTCGGT 29265
 QY 438 erGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnG 458
 DB 29264 CGGGCCAGCAATTCGCCGAAAGCGGGTGTGTAAGGATTCGAAAGGTTCGACATAGT 29205
 QY 458 lNLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnA 478
 DB 29204 TTTTGGAAATGTCAACCTGTCTCAGTAAGTGAATGGTGAACATGATTGTCCCCAGC 29145
 QY 478 tGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeuGlnLysAla 498
 DB 29144 GTGCGTATGAGTTTAAATTCAAAGCAATTCACACGAGTGACAACTGCTCGGCACAGCAG 29085

US-10-009-823A-1 (1-502) x AAT67783 (1-816)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 28 ATGCTCGCTCTCTATAGTGCACCTTCAGGATGCTCGCCCAACAAACGCAATTCAC 87
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 88 ACCACTTCAAAACAACATCGCAATGTCAATACACCGGGTTTAAAAAATCTCGCGGAT 147
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 DB 148 TTTAACGACTTGTTTTACCAAGCATGCAATACCCCGCACCAACACAAAGCAACACGAT 207
 QY 61 -----GlnAlaGlyMetGlyAlaGlnValGlnValGlySerValArgThrIle 74
 DB 208 TTATGCCAGATGCGATCGAAGTGGCTTGGCTAGCGCTAGCTAGCGATTACCAAAATG 267
 QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys 94
 DB 268 TTTTCGCAAGGCAGCCCTTAAGAAACGAGAGATAATTTAGATATTGCTATTACAGTAAA 327
 QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
 DB 328 GCGTTTTTCAAGTCCAGCTTCTGATGGCACTACCGCTTACACAGAGGCGGGAATTC 387
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 DB 388 AAGCTAGACGAGCAGGCAATCTTGTAAACAGCGAGGCTATCTCTCATC----- 438
 QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 DB 439 -----CCTCAATC-----ACTTTA----- 453
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 DB 454 -----CCGAGACACACCGCAA-----GTGAAT 477
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
 DB 478 ATCGGT----- 483
 QY 194 SerTyrLysAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 DB 483 ----- 483
 QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
 DB 484 -----GTGATGGGACG 495
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 DB 496 GTGACGCGTACT----- 507
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 507 ----- 507
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 DB 507 ----- 507
 QY 294 AspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
 DB 507 ----- 507
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 DB 507 ----- 507
 QY 334 MetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353

QY 498 euGluLeuLysArg 502

DB 29084 TGGCGCTCAGCGT 29071

RESULT 29

AA67783

ID AAT67783 standard; DNA; 816 BP.

AC AAT67783;

XX 29-JUL-1997 (first entry)

DE H. pylori flagella associated protein ORF 29298130.aa.

XX Vaccine; prevention; treatment; infection; identification;

XX binding compound; bacterium; life cycle; activator; bacteria; inhibitor;

KW duodenal ulcer disease; chronic gastritis; diagnosis; flagella;

KW associated; flagellum; ds.

XX Helicobacter pylori.

OS

XX

XX Key Location/Qualifiers

FF CDS 1..816

FT /*tag= a

XX WO9640893-AL.

XX 19-DEC-1996. 96WO-US009122.

XX 06-JUN-1996; 96WO-US009122.

XX 07-JUN-1995; 95US-00487032.

PR 01-APR-1996; 96US-00630405.

XX (ASTR) ASTRA AB.

XX Smith D, Berglindh OT, Mellgaerd BL;

PI WPI; 1997-052306/05.

DR P-PSDB; AAW20373.

XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -

PT useful for vaccines to treat or prevent H. pylori infection, and to

PT detect Helicobacter.

XX Claim 1; Page; 1481pp; English.

XX The present sequence encodes a Helicobacter pylori flagella associated

CC protein. The protein may be used in a vaccine to prevent or treat H.

CC pylori infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial DNA.

CC The sequences were analysed for ORF of at least 180 nucleotides, and the

CC predicted coding regions defined by computer evaluation. To identify

CC likely H. pylori antigens for vaccine development, the amino acid

CC sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts. Note: This DNA sequence is not

CC reproduced in the specification and has been derived from the related

CC specification, WO9719098

XX Sequence 816 BP; 237 A; 199 C; 183 G; 197 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4.65e-14 Length: 816

Score: 288.00 Matches: 107

Percent Similarity: 28.68% Conservatives: 39

Best Local Similarity: 21.02% Mismatches: 109

Query Match: 11.14% Indels: 254

Db 507 ----- 507
Qy 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer 373
Db 507 ----- 507
Qy 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 507 ----- 507
Qy 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 508 ----- 546
Qy 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 547 ACTTGGCTAATTTGTCAATCCGGCGGCTTCATTCTATGGGGATATTTGTTTTC 606
Qy 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 607 ATCACCACGCTAGCGCGATGCGATTGTGGCAACCCGGATTCTCAAGGCTTAGGCAAG 666
Qy 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 667 TTAAGGCAAGCTTTTGGAGCTTAGTAACGTGAGATTGGTAGAAGAAATGACAGATCTA 726
Qy 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 727 ATCACCCTCAAGGGCTTATGAAGCAATTTAAAGCAATTCACACCGCTGATGCCATG 786
Qy 494 LeuGlnLysAlaLeuGluLeuLysArg 502
Db 787 CTCCTCAACAGTCATTCCTCAACACG 813

RESULT 30

AAAT77463
ID AAAT77463 standard; DNA; 816 BP.

AC AAAT77463;

DT 11-AUG-1997 (first entry)

DE H. pylori flagella associated protein ORF 29298130.aa.

KW Chronic gastritis; duodenal ulcer disease; activator; inhibitor;
KW bacterial life cycle; vaccine; immunisation; detection; antisense;
KW inhibition; flagella; flagellum; basal body; rod; ds.

OS Helicobacter pylori.

PH Key Location/Qualifiers
CDS 1..816
/*tag= a

FN WO9719098-A1.

PD 29-MAY-1997.

PF 15-NOV-1996; 96WO-US018542.

PR 17-NOV-1995; 95US-00561469.

PA (ASTR) ASTRA AB.

PI Smith DH;

DR WPI; 1997-298052/27.

DR P-PSDB; AAW24645.

PT Helicobacter pylori nucleic acid sequences and related proteins - used
PT for diagnostics and therapeutics.

PS Claim 1; Page 105; 235pp; English.

XX The present sequence encodes a Helicobacter pylori flagella associated
CC protein, which was found to be homologous to flagellar basal body rod
CC protein following BLAST protein analysis. H. pylori has been strongly
CC linked to chronic gastritis and duodenal ulcer disease. The nucleic acid
CC sequences of the invention are used to evaluate compounds, especially
CC activators or inhibitors of bacterial life cycle, for the ability to bind
CC an H. pylori nucleic acid sequence. The nucleic acid sequences, and
CC corresponding proteins, are also useful for generating vaccines for
CC immunising subjects against H. pylori or for use in detecting the
CC presence of Helicobacter species in a sample. Antisense nucleic acid
CC sequences of these sequences are used to inhibit expression of a gene
CC from Helicobacter species. H. pylori whole genomic DNA was isolated and
CC nebulised to a median size of 2000 bp. Purified DNA fragments were blunt-
CC ended and ligated to unique BstXI-linker adapters in 100-1000 fold molar
CC excess. These linkers are complementary to the BstXI-cut pMPX vectors,
CC while the overhang is not self-complementary. Therefore the linkers will
CC not concatemerise nor will the cut vector re-ligate itself easily. The
CC linker-adaptor inserts were ligated to each of the 20 pMPX vectors to
CC construct a series of shotgun subclone libraries. The purified DNA
CC samples were then sequenced. Note: The ORF/protein reference number for
CC this sequence was obtained from the related specification, WO9640893

XX Sequence 816 BP; 237 A; 199 C; 183 G; 157 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4.65e-14 Length: 816
Score: 288.00 Matches: 107
Percent Similarity: 28.68% Conservative: 39
Best Local Similarity: 21.02% Mismatches: 109
Query Match: 11.14% Indels: 254
DB: 2 Gaps: 9

US-10-009-823A-1 (1-502) x AAT77463 (1-816)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 28 ATGTCCTGCTCTCTCTATAGTCCACTTCAGGATGTCGCCCAACACGACATTGAC 87
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 88 ACCACTTCAAAACAACATCGCCCAATGTCAATACCAACCGGGTTTAAAAAATCTCGCGGAT 147
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 148 TTTAAGCACTGTTTTTACCAAGCGATGCAATACGCCGACCAACACAGCAACACCACT 207
Qy 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
Db 208 TTATCGCCAGATGCATGGAAGTGGGCTTGGCTAGCCCTAGTGGGATTACCAAAATG 267
Qy 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
Db 268 TTTTCGCAAGCAGCCCTTAAGAAACCGGAGAATAATTTAGATATTTCTATACGGTAAA 327
Qy 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
Db 328 GGCCTTTTTCAGTCCAGCTTCCTGATGGCACTACCGCTTACACAGGAGCGGGAATTC 387
Qy 114 ArgPheThrGlnAspGlyPheLeuAsnAspSerSerGlyPheThrLeuMetGlySerArg 133
Db 388 AAGCTAGCAGCAGCGGCAATCTTGTAAACAGCGAGGGCTATCTCTCATC----- 438
Qy 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 439 -----CCTCAATC-----ACITTA----- 453
Qy 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 454 -----CCGAAAGACACACCGCAA-----GTGAAT 477
Qy 174 LeuGluAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193

478	ATCGGT	---	483
194	SerTrpLysGlyAsnGlyThrProPheSerThrSerAsnTyrSerTyrAlaGlnPro	213	
483	---	483	
214	MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla	233	
484	---	495	
234	ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp	253	
496	GTGAGCGTGACT	---	507
254	GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe	273	
507	---	507	
274	SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys	293	
507	---	507	
294	AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn	313	
507	---	507	
314	PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnAsn	333	
507	---	507	
334	MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro	353	
507	---	507	
354	SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSer	373	
507	---	507	
374	ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr	393	
507	---	507	
394	SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle	413	
508	---	546	
414	ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer	433	
547	ACTTTGGCTAATTTTGTAATCCGCGGGGCTTCATTCTATGGGGGATTAATTGTTTCC	606	
434	AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys	453	
607	ATCACCAACGCTAGCGGGGATGCGATTGTGGCAACCCGGATTCTCAAGGCTTAGGCAAG	666	
454	LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet	473	
667	TTAAGCGAAGGCTTTTTCGAGCTTAGTAACGTGAGATTGGTAGAAGAAATGACAGATTCTA	726	
474	IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet	493	
727	ATCACCGCTCAAGGGCTTATGAAGCCAAATCTTAAAGCATTCAACCCGCTGATGCATG	786	
494	LeuGlnLysAlaLeuGluLeuLysArg	502	
787	CTCCAAACAGTCAATTCCCTCAAAACGC	813	

RESULT 31
 AA768116
 ID AA768116 standard; DNA; 837 BP.
 XX
 AC
 AA768116;
 XX

DT	16-JUL-1997	(first entry)
XX	H. pylori	flagella-associated membrane protein ORF 129e20305orf11.
XX		Cytoplasmic; vaccine; prevention; treatment; infection; identification;
XX		binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW		duodenal ulcer disease; chronic gastritis; diagnosis; envelope; membrane;
KW		flagella; ds.
XX		
XX	Helicobacter pylori.	
XX		
XX	Key	Location/Qualifiers
XX	CDS	1..837
FT		/*tag= a
FT		/note= "no stop codon given in sequence"
FT		
XX	W09640893-Al.	
PN		
XX	19-DEC-1996.	
PD		
PD	06-JUN-1996;	96WO-US009122.
PP		
XX	07-JUN-1995;	95US-00487032.
XX	01-APR-1996;	96US-00630405.
PR		
XX	(ASTR)	ASTRA AB.
XX		
XX	Smith D, Berglindh OT, Mellgaerd BL;	
PI		
XX	WPI; 1997-0523206/05.	
DR	P-PSDB; AAW20863.	
DR		
XX		
XX	Helicobacter pylori	nucleic acid sequences and related polypeptide(s) -
PT		useful for vaccines to treat or prevent H. pylori infection, and to
PT		detect Helicobacter.
PT		
XX		
XX	Claim 1; Page 909; 1481pp; English.	
PS		
XX		
XX	The present sequence encodes a Helicobacter pylori flagella-associated	
CC	membrane protein. The protein may be used in a vaccine to prevent or	
CC	treat H. pylori infection or to identify H. pylori polypeptide binding	
CC	compounds, useful as potential H. pylori life cycle activators or	
CC	inhibitors. The genomic sequence of H. pylori (ATCC 55679) was determined	
CC	from overlapping contigs generated by mechanically shearing the bacterial	
CC	DNA. The sequences were analysed for ORF of at least 180 nucleotides, and	
CC	the predicted coding regions defined by computer evaluation. To identify	
CC	likely H. pylori antigens for vaccine development, the amino acid	
CC	sequences predicted from various ORF were analysed for significant	
CC	homology to other known or exported membrane proteins. Having identified	
CC	and determined the sequences of interest, particular regions can be	
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide	
CC	production, e.g. in E. coli hosts	
XX		
SO	Sequence 837 BP; 240 A; 201 C; 186 G; 210 T; 0 U; 0 Other;	

Alignment Scores:

Alignment Scores:		
Pred. No.:	4.8e-14	Length: 837
Score:	288.00	Matches: 107
Percent Similarity:	28.6%	Conservative: 39
Best Local Similarity:	21.0%	Mismatches: 109
Query Match:	11.1%	Indels: 254
DR:	2	Gaps: 9

US-10-009-823A-1 (1-502) x AAT68116 (1-837)

QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 172 TTTAAACGACTTGTCTTACAGCGATGCAATACGCGGCACCAACAGCAACACGACT 231
QY 61 -----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
Db 232 TTATCGCCAGATGGCATGGAAGTGGCTTGGCTACGCCCTAGTCGATTACCAAAATG 291
QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
Db 292 TTTTCGCAAGGCGAGCCCTAAAGAAACGGAGAGATAATTAGATATTGCTATTACAGGTAA 351
QY 95 GlyPhePheGlnValThrLeuGluAsp--LysValHisTyrThrArgAlaGlyAsnPhe 113
Db 352 GGCCTTTTCAAGTCCAGCTTCTGTAGCCACTACCGCTTACACAGAGCGGAATTC 411
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 412 AAGCTAGACGACGAGCGCAATCTTGTAACAGCGAGGCTATCTCTCATC----- 462
QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 463 -----CCTCAATC-----ACTTTA----- 477
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 478 -----CCGACAGACACACGCA-----GTCAAT 501
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
Db 502 ATCGGT----- 507
QY 194 SerTyrLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
Db 507 ----- 507
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
Db 508 -----GTGATGGCAGC 519
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 520 GTGACCGTACT----- 531
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 531 ----- 531
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 531 ----- 531
QY 294 AspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 531 ----- 531
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyLysSerGlnGlnAsn 333
Db 531 ----- 531
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 531 ----- 531
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 531 ----- 531
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 531 ----- 531
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413

Db 532 -----CAAGCTTGCACACGACTTCTTAACGTGATCGGG-----CAAAATC 570
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSer 433
Db 571 ACTTGGCTAAATTTGTCAATCCGGGGGCTTCAATTCATGCGGGGATAATTTGTTTCC 630
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 631 ATCACCACGCTAGCGCGATCGGATTGCGCAACCCGGATTCTCAAGGCTTAGGCAAG 690
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 691 TTAAGCAAGGCTTTTGGAGCTTAGTACGTGAGATTGGTAGAAGAATGACATCTA 750
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 751 ATCAGCGCTCAAAGGCTTATGAAGCCAAATCTAAAGCAATTCAAACCGCTGATGCCATG 810
QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502
Db 811 CTCCAACAGTCAATTCCTCCTCAACGC 837
RESULT 32
ADF03653
ID ADF03653 standard; DNA; 798 BP.
XX ADF03653;
AC ADF03653;
DT 12-FEB-2004 (first entry)
XX
XX Bacterial polynucleotide #3938.
XX
XX Proteus mirabilis infection; bacterial infection; antibacterial;
KW immunostimulant; gene; ds.
XX
XX Proteus mirabilis.
XX
XX US6605709-B1.
XX
XX 12-AUG-2003.
XX
XX 05-APR-2000; 2000US-00543681.
XX
XX 09-APR-1999; 99US-0128706P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX
XX WPI; 2003-895291/82.
DR P-PSDB; ADF07825.
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
XX
XX Disclosure; SEQ ID NO 3938; 870pp; English.
XX
XX The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polynucleotide of the invention.
XX

SQ Sequence 798 BP; 233 A; 163 C; 189 G; 213 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,55e-14 Length: 798
 Score: 286.00 Matches: 100
 Percent Similarity: 30.57% Conservative: 55
 Best Local Similarity: 19.72% Mismatches: 98
 Query Match: 11.06% Indels: 254
 DB: 10 Gaps: 9

US-10-009-823A-1 (1-502) x ADF03653 (1-798)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 16 ATGATCGGTTCTTATGGATTGCTAAACACAGGGTGGATGCACAACTAACATGAT 75
 QY 21 ThrValSerAsnAlaAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 76 GTGATTCCACAACTCCCAACGTCAGCACAATGGTTTAAAGCCAGCGTGGCGTT 135
 QY 41 PheGlnAspLeuPheSerGlnAspLeu-----AlaIleGlySerThrGlySer 56
 DB 136 TTGAGATTACTCTATCAACTATTCGTCACCGGGAGCGATGACATCCGAGACAG 195
 QY 57 GlnGlyProAsn-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 DB 196 AATGCGCCTCTCGTTTACAAATTGGTACTGTGTTCGCCAGTGGCGGACAGACGTTT 255
 QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyLys 94
 DB 256 CATAGCCAGGTAATTTAGCCCAACTATGGTACCGGTGATGTTGCTATCAAGGGCAA 315
 QY 95 GlyPhePheGlnValThrLeuGluAspLysValHis---TyrThrArgAlaGlyAsnPhe 113
 DB 316 GGTITTTTCCATGTTCAATTTACCTGATGGTACGATGCTTATCTCGTGTGCTCTTT 375
 QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
 DB 376 CAAATGGACCAAAATGGCACTAGTACCTCCAGTGGCTTTCAAAATCGTCCAGCG 432
 QY 134 IleSerAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
 DB 433 -----ATTATTTTCCAGAAACCGCTAAAGGTGATGTA----- 468
 QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
 DB 468 ----- 468
 QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuGlu 193
 DB 468 ----- 468
 QY 194 SerTrpLysGlyAsnGlyThrProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
 DB 468 ----- 468
 QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
 DB 469 -----GGTGGTGGTGGTATGTCAGTGTGAGATTGAGATCG 507
 QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
 DB 508 CCTGCA----- 513
 QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
 DB 513 ----- 513
 QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
 DB 513 ----- 513
 QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313

DB 514 -----CCTCAA----- 519
 QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
 DB 519 ----- 519
 QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
 DB 520 -----CAAGTAGGCGCAATTA--- 534
 QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
 DB 534 ----- 534
 QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
 DB 535 -----ACCTCACC 543
 QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
 DB 544 ACA----- 546
 QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
 DB 547 -----TTTATTATGATAGCGGTAGAAAGTGTGGGAAATCTGTACTTA 594
 QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
 DB 595 GAAACAGCAGCTCCGGAGCACCCACTGAGAAATCGCGGTATTAAACGTCGCGGCTTG 654
 QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
 DB 655 TTATATCAGGATATGTTGAAACCTCTAACGTTAATGTCCGGGAGAAATGTCATATG 714
 QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
 DB 715 ATCCAAACTCAGCGTGTATTGAAATTAATAGTAAGCGATTTCACACTTCTGATCAGATG 774
 QY 494 LeuGlnLysAlaLeuGluLeu 500
 DB 775 TTACAGAAACTCAGCACTC 795

RESULT 33

AZ288552
 ID AZ288552 standard; DNA; 1800 BP.
 AC AZ288552;
 DT 04-MAY-2000 (first entry)
 XX C. jejuni flgP and flgG DNA.
 DE Basal body rod protein; flgP; flgG; flagellum; vaccine; immunogen;
 KW pathogenic bacteria; detection; antibacterial; ss.
 XX Campylobacter jejuni.
 OS
 PH Key Location/Qualifiers
 FT CDS 101..913
 FT /*tag= a
 FT /product= "flgP"
 FT 942..1733
 FT /*tag= b
 FT /product= "flgG"
 XX
 PN US6020125-A.
 XX
 PD 01-FEB-2000.
 XX
 PF 07-JUN-1995; 95US-00483857.
 XX
 PF 08-MAY-1995; 95US-00436748.
 PR

XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Louie H, Chan VL;
 XX DR WPI: 2000-146975/13.
 XX DR P-PSDB; AAY51376, AAY51377.
 XX
 PT Recombinant basal body rod protein producible by a transformed host for
 PT use in immunological vaccine compositions for in vivo administration to
 PT protect against diseases caused by bacterial pathogens.
 XX
 PS Claim 1a; Col 15-20; 28pp; English.
 XX
 CC This invention describes a novel recombinant basal body rod protein (I)
 CC producible by a transformed host containing an expression vector
 CC comprising a nucleic acid selected from; (a) an entire nucleic acid
 CC sequence of 1800 base pairs (bp), or the nucleic acid sequence of the
 CC flgF gene having 810 bp, both given in the specification; (b) a nucleic
 CC acid encoding the amino acid sequence of the flgF protein, a 270 residue
 CC sequence, given in the specification; (c) a nucleic acid sequence
 CC encoding a functional flgF basal body rod protein of a flagellum of a
 CC strain of *Campylobacter*; or (d) an immunogenic fragment of an flgF
 CC protein of (a), (b) or (c), and expression means operatively coupled to
 CC the nucleic acid molecule for expression by the host of a basal body rod
 CC protein of a flagellum of a strain of *Campylobacter*. (I) is useful in
 CC immunological vaccine compositions for in vivo administration to protect
 CC against diseases caused by bacterial pathogens that produce basal body
 CC rod proteins. (I) are also useful as immunogens, as antigens in
 CC immunoassays, or for procedures for the detection of antibacterial,
 CC *Campylobacter*, basal body rod protein and/or peptide antibodies. The
 CC product of the invention has immunostimulatory activity. This sequence
 CC encodes the *Campylobacter* jejuni flgF and flgG proteins described in the
 CC method of the invention

SQ Sequence 1800 BP; 636 A; 251 C; 363 G; 550 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.18e-13 Length: 1800
 Score: 281.50 Matches: 133
 Percent Similarity: 38.41% Conservative: 94
 Best Local Similarity: 22.50% Mismatches: 228
 Query Match: 10.89% Indels: 136
 DB: 3 Gaps: 22

US-10-009-823A-1 (1-502) x AAZ88552 (1-1800)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 DB 101 ATGCAAAATGGATATTATCAAGCAACTGCGGGAATGGTAATCACTCAGTTTATAAATCTGAT 160
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 DB 161 GTGATTACTAATATCTTGCCATATCAATACAGTGGATATAAAGAGATGATGGTT 220
 QY 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlaIleGlySer--- 53
 DB 221 ATTCAGATTTTTAAAGGATTTTTTAAAGAACTCAGGATGATGCTATAGAAAATCAC 280
 QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly---AlaGlnValGlySer 70
 DB 281 ACAAGATGCAATCTGTTTGTAAATPACTACATAGATGGAATCCCAAGTTTCTCAA 340
 QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
 DB 341 GAATATACGGATTTTAGCTAGGTTCTTTAAAGGCCACAAACAATCTTGGATTGGCA 400
 QY 91 IleGlyGlyLysGlyPhePheGlnVal-----ThrLeuGluAspLysValHisTyrThr 108
 DB 401 ATGACTAGAGAAGATGCTTTTATTGTCACACCAAGATCGAGAAGTAAGATTAACC 460
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 DB 1409 AACCGTTTCTGTAATGCTTACCAGGGGAGCAACAAGAACTCAATTT-----GGCCA 1459

DB 461 AAAGATGGAAATTTTCAACTTTGATGATGAGGGTTATTGGTAAATAAGCAAGGATACAAG 520
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsn----- 139
 DB 521 GTATTAAAGTAGTATATTATTAATCTCAGATGCTGGCATACGATTCCTTAATAGT 580
 QY 139 ----- 139
 DB 581 GCTGTTCAAAATTAGCGTTGATAAAAAACGAAGCATTGAAGTTGATGGAGCTCAAAATGCA 640
 QY 140 -----IleLysLysGlnThrLeuGluProIleGlnLeuAspPheAsn----- 153
 DB 641 AGATTATTGTCACCAAGTAGATATATAGAGCTTTTGCAAAAGATGGGGATATATGTC 700
 QY 154 -----AspProThrValAlaLysSerProLalysThrSerThrAlaLeuAsnAla 170
 DB 701 TATAAATAGATGATCTAACCCGATTATAGAGATTGAAAAAATCCCAATGCTATTTCGCCAA 760
 QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSer-----GluAla 185
 DB 761 GGTGTTTCTCAGGGATCAAAATGTTAATCCAGTTACTGAAATGGTAGGACTGATTGAAGCA 820
 QY 186 AsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProIleSerThr 205
 DB 821 AAC-----AGAATCGTAGAAATGATCAAAAAAGTTATGACAGCTCATATGGATGAC 871
 QY 206 SerAsnTyrSerTyrAlaGlnProMet----- 214
 DB 872 TTAATCAAGAGCTATCAATAAGCTTGCAGCTGTTTAATAATTAAATAAATAAAAA 931
 QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
 DB 932 AGGATTAATAATGATGAGATCACTTCATAC-----TGCTGTACAGGAATGGTAGCGCA 985
 QY 235 SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
 DB 986 GCAAAACACAAATTTGATGTTTACTTCAATAACATCGCCAATGTTAATACAGCA-----GG 1039
 QY 254 ySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeu-----MetSerGly 269
 DB 1040 TTTTAAAGAAAGTCCGCGAGATTTGCTGATCTTATGATCATAGTTATGAAGTAGCAGG 1099
 QY 269 yThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly 289
 DB 1100 AACTTCAACT-----TCAGCTACTACTCTCTTCTCTCTCGGG 1135
 QY 289 ySerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGly 309
 DB 1136 TATAGAA----- 1142
 QY 309 nPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu----- 324
 DB 1143 -----GTGGGTGGGTGTGCGTCCCAACAGCGGTAACTAAAGTTTTCAC 1186
 QY 325 AspPheGlyIleLysSerGlnGlnAsnMetThrAlaGlyAlaProAlaSerAlaAla 344
 DB 1187 TCAAGGAAATTTTAAATCAACAAGTACT-----GATGGTCTTGATATGCTATGTCAGG 1240
 QY 344 aIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAs 364
 DB 1241 TAATGGGTTTTTCAATAACAACCTTCTGTGGCATTAGGATATAGTAATGGCA 1300
 QY 364 nSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyr----- 382
 DB 1301 ATTTACAAAAGATAATGAAGGTAATATTGTA-----AATTTCAGATGTTTATAGACT 1351
 QY 383 -----ProGlnGlyAspLeuValAspValThrIleThrSerGluGly 396
 DB 1352 TTTACCTGAATGACATACCTGAAGGC-----GCACAGCAATTAATGTTGCTACAGATGG 1408
 QY 396 y-----LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAs 412
 DB 1409 AACCGTTTCTGTAATGCTTACCAGGGGAGCAACAAGAACTCAATTT-----GGCCA 1459

CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS84197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1398 BP; 292 A; 401 C; 346 G; 359 T; 0 U; 0 Other;

Alignment Scores: Length: 1398
Pred. No.: 4,85e-13 Matches: 77
Score: 279.00 Conservative: 30
Percent Similarity: 49.77% Mismatches: 92
Best Local Similarity: 35.81% Indels: 16
Query Match: 10.73% Gaps: 5
DB:

US-10-009-823A-1 (1-502) x AAS88418 (1-1398)
QY 285 PheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304
Db 1386 TTTGTGAAGACCGGGAT-----AATAACTGGCAGGTCTACACCCAGAT 1342
QY 305 AsnGlyLeuProGlnPheSerAlaAsnPhValGlyAlaGlylleGlnProLeuThrLeu 324
Db 1341 AGCAGTGTATCCAAACAGCATTTGCCAAG-----ACAGCGACACACTG 1300
QY 325 AspPheGlylleLysSerGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAla 344
Db 1299 GAATTT-----AATGCTAATGGCACATTAGTGATGGTGCGAATAAATATCGCA 1246
QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
Db 1245 ACCGGCGCAATTAACGGTGCAGAACCCGCCAGTTTAGTCTG-----AGTTCTCTCAAC 1192
QY 365 SerThrAlaArgAsn---GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrPro 383
Db 1191 TCCATGCGCAAAATACCGCGCTTAACAATATTGTGGCAACCCAGACGGGTACAAA 1132
QY 384 GlnGlyAspLeuValAspValThrIleThrSerGlyLysLeuGlnGlyLysTyrSer 403
Db 1131 CCGGGCGATCTGGTGAGTTATCAATCAATGATGACGGTACGGTGTGCGCAACTATTTC 1072
QY 404 AsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGly 423
Db 1071 AACGAACAAACCCCACTCTGGGCGAGATTGTACTGGCGAACTTTGCCAAACCAACGAGT 1012
QY 424 LeuArgArgGlyLysAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPhe 443
Db 1011 CTGGCATCCGAAGCGCACACAGCTGTGCTGGAGCGCAATCTTCTGGCGTGGCGCTGTG 952
QY 444 GlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsn 463
Db 951 GGGACAGCGCGGACCGGAAACTTTGGCACCTTGACCAACCGGTGCGCTGGAACGGTCCAAC 892
QY 464 ValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsn 483
Db 891 GTCGATCTCAGPACAGAACTGTGTCAATATGATGCTTGGCCAGCGTAATATCAGTCTAAC 832
QY 484 SerLysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeu 498
Db 831 GCCCAGACCATCAAAACCCAGGACCCAGATCTCTCAACACGCTGCTC 787

RESULT 35
AAV58977
ID AAV58977 standard; DNA; 1800 BP.
XX AAV58977;
AC AAV58977;
XX 05-JAN-1999 (first entry)
DT
XX FlgFG operon.
DE

QY 412 nIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGlyLysAsnAsnHisTyr 432
Db 1460 AGTGGAGTAGTTCAGTTTATAAATCCAGCGGTCTTCATTTATGGTGTATAATCTTTA 1519
QY 432 rSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrG1 452
Db 1520 TCTTGAACAGGAGCAAGTGGTGCACCTTTTCGGGTATAGCAGGACAGATGGGCTTGG 1579
QY 452 yLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAs 472
Db 1580 AACATTAACACATGGTTTATAGACTTATGTTTCAGCTTGTGAAGAAATGACAGA 1639
QY 472 nMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspTh 492
Db 1640 TCTTATCAGACGACAAAGAGCTTATGACGCGGTTCATAGCGCATTTACAAACAGTATGA 1699
QY 492 rMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 1700 TAGCTAGGAATTGTAATCAGCTTAAGCGA 1730

RESULT 34
AAS88418/c
ID AAS88418 standard; cDNA; 1398 BP.
XX
AC AAS88418;
XX
DT 13-FEB-2002. (first entry)
XX
DE DNA encoding novel human diagnostic protein #24222.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
PR
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG24231.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 24222; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations

XX flgFG operon; flgF protein; flgG protein; basal body rod protein;
 KW flagellum; secretory diarrhoea; enteritis; vaccine; therapy; ds.
 XX

OS Campylobacter jejuni.

XX Key Location/Qualifiers

XX CDS 101..913

XX /tag= a

XX /product= "flgF"

XX CDS 942..1733

XX /tag= b

XX /product= "flgG"

XX US827654-A.

XX 27-OCT-1998.

XX 08-MAY-1995; 95US-00436748.

XX 08-MAY-1995; 95US-00436748.

XX (UTOR) UNIV TORONTO.

XX Louie H, Chan VL;

XX WPI; 1998-593983/50.

XX P-PSDB; AAW73072, AAW73073.

XX DNA encoding Campylobacter flagellum basal body rod proteins - useful for
 DR recombinant production of the proteins for use as vaccines against the
 DR bacterium, and for its detection, additionally with antibodies raised
 PT with the protein.
 XX

PS Claim 1; Fig 1; 27pp; English.

XX This sequence represents the Campylobacter flgFG operon of the invention.
 CC The encoded flgF and flgG proteins are basal body rod proteins of the
 CC flagellum of the Campylobacter. The nucleic acid and host cell containing
 CC it, are useful for the recombinant production of Campylobacter,
 CC especially C. jejuni, basal body rod proteins. This bacterium is the
 CC cause of secretory diarrhoea and enteritis. As such, the peptides
 CC produced can be used to raise antibodies, which in turn can be used to
 CC detect the presence of the organism diagnosis of the conditions.
 CC Additionally, the peptides, and specifically the live vectors (e.g. pox-
 CC or vaccinia virus) can be used as vaccines against the bacterium, and the
 CC antibodies can be used for passive immunisation. The nucleic acids can
 CC also be used to detect the presence of the bacterium

XX SQ Sequence 1800 BP; 637 A; 250 C; 364 G; 549 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.79e-13 Length: 1800
 Score: 277.50 Matches: 132
 Percent Similarity: 37.94% Conservative: 96
 Best Local Similarity: 21.96% Mismatches: 156
 Query Match: 10.73% Indels: 24
 DB: Gaps: 2

US-10-009-823a-1 (1-502) x AAV58977 (1-1800)

QY 1 MetMetGlySerLeuPheGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 101 ATGCAAAATCGATATTATCAAGCACTGGCGGAATGTAACAGTTTAACTGAT 160
 QY 21 ThrValSerAsnLeuAlaAsnAlaThrGlyTyrGlnGlnValVal 40
 Db 161 GTGATTACTAATAATCTTGCATATATCAATACAAAGTGGATATAAAGATGATGTGTT 220
 QY 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlaIleGlySer--- 53
 Db 221 ATTGCAGATTTTAAAGGATTTTAAAGAACTCAGGATGATGCTTATAGAAATCAC 280

QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly-----AlaGlnValGlySer 70
 Db 281 ACAAGAGATGCATCTCGTTTGTAAATACTACAATAGATGGAATCCCAAGTTTCTCAA 340
 QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
 Db 341 GAATATACGATTTTAGCTAGCTTCTTTAAAGGCCACAAACAATCTTTGGATTGGCA 400
 QY 91 IleGlyGlyLysGlyPhePheGlnVal-----ThrLeuGluAspLysValHisTyrThr 108
 Db 401 ATGACTAGAGAAGATGCTTTTATTGTTGTCAGACCAAGAGTGGAGAGTAAAGATTACC 460
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 Db 461 AAAGATGGAAATTTTCAACTGATGATGAGGGTATTGTTGTAATAAGCAAGGATACAG 520
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsn----- 139
 Db 521 GTATTAAAGTAGTGATTTATTTAATAATCTCAGAACTGTCGATACGATTCCTAATAGT 580
 QY 139 ----- 139
 Db 581 GCTGTTCAATTAGCGTTGATATAAAACGGAAGCATTTGAAGTTGATGGAGCTCAAAATGCA 640
 QY 140 -----IleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn----- 153
 Db 641 AGATTATTGTTAGCACAAAGTAGATGATATAAGAGCTTTGCAAAAGATGGGATAATGTC 700
 QY 154 -----AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
 Db 701 TATAAATAGATGATCTAACCCGTTATTAGATTGTAATAAACTCCCAATGCTATTTCGCCAA 760
 QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSer-----GluAla 185
 Db 761 GGTTTTCTCAGGGATCAATGTTAATCCAGTTACTGAAATGGTAGGACTGATTGAAGCA 820
 QY 186 AsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThr 205
 Db 821 AAC-----AGAATGGTAGAAATGTATCAAAAGTTATGACAGCTCATATGGATGAC 871
 QY 206 SerAsnTyrSerTyrAlaGlnProMet----- 214
 Db 872 TTAATCAAGAGCTATCAATAGCTTGCAGCTTTAAATAATTTAAATAAATAAATAA 931
 QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
 Db 932 AGGATTAATAATGATGATGATCACTTCATAC-----TGCTGCTACAGGAATGGTAGCGCA 985
 QY 235 SerSerThrGly-SerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspG1 254
 Db 986 CCAACCAAAATTTGATGTTTCTTCAATAACATCGCAATGTTAATACAGCA-----GG 1039
 QY 254 YSerAlaAlaSerGlyThrAspSerAlaGlyLeuLeu-----MetSerG1 269
 Db 1040 TTTTAAAGAAAAGTCGCGCAGAAATTTGCTGATCTTATGATCAAGTTATGAAGTATGCGAG 1099
 QY 269 YThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrG1 289
 Db 1100 AACTTCAACT-----TCAGTACTACTCTCTTCTCTCTGGG 1135
 QY 289 YSerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProG1 309
 Db 1136 TATAGAA----- 1142
 QY 309 nPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu----- 324
 Db 1143 -----GTGGTGTGGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 1186
 QY 325 -AspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAla 344
 Db 1187 TGAAGGAATTTAAATCAACAAGTACT-----GATGGTCTTGATATGCTATTGTCAGG 1240
 QY 344 alleGly-----ThrAspIleGlyLysLeuProSe 354

Db 1241 TAATGGGTTTTCATTAATACAACTTCCTGATGGCACTATAGAAATGGGCAATTTACAAA 1300
 QY 354 rMetMetProLleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerTh 374
 Db 1301 GATATG-----AAGGATAAGCAAGGTAATATTGTA----- 1331
 QY 374 rArgArgTyrSerGlnAspGlyTyr-----ProGlnGlyAs 386
 Db 1332 -----AATTCAGATGGTTATAGACTTTTACCTGAAATGACAATACCTGAAGGC-- 1379
 QY 386 pLeuValAspValThrIleThrSerGluGly-----LysLeuGlnGlyLysTy 402
 Db 1380 -GCACAGCAATAATGTTGTACAGATGGACCGTTTCTGTATGTACAGGGAGCA 1438
 QY 402 rSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAs 422
 Db 1439 ACAAGAACTCAAAAT-----GGCAAGTGGAGCTAGTTCAAGTTTATAATCCAGC 1489
 QY 422 pGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProG1 442
 Db 1490 GGGTCTTCATTCATGGGTGATTAATCTTAACTTGAACAGAGCAAGTGGTGCACCTGT 1549
 QY 442 uPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSe 462
 Db 1550 TCGGGTATAGCAGGACCAAGATGGGCTTGGACCAATAGACATGGATTATAGAACTTAG 1609
 QY 462 rAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMe 482
 Db 1610 TAATGTTCACTGTGTGAGAAATACAGATCTTATACAGACACAGAGCTTATGAGC 1669
 QY 482 tAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysAr 502
 Db 1670 GGGTCTTAAGGCCATACACAAAGTGATGATGCTAGGAATTTGAATCAGCTTAAGCG 1729
 QY 502 g 502
 Db 1730 A 1730

RESULT 36
 AAZ88554
 ID AAZ88554 standard; DNA; 789 BP.
 XX
 AC AAZ88554;
 XX
 DT 04-MAY-2000 (first entry)
 XX
 DE C. jejuni flgG DNA.
 XX
 KW Basal body rod protein; flgG; flagellum; vaccine; immunogen;
 KW pathogenic bacteria; detection; antibacterial; ss.
 XX
 OS Campylobacter jejuni.
 XX
 PN US6020125-A.
 XX
 PD 01-FEB-2000.
 XX
 PF 07-JUN-1995; 95US-00483857.
 XX
 PR 08-MAY-1995; 95US-00436748.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Louie H, Chan VL;
 XX
 DR WPI; 2000-146875/13.
 DR P-PSDB; AAZ51377.
 XX
 PT Recombinant basal body rod protein producible by a transformed host for
 PT use in immunological vaccine compositions for in vivo administration to
 PT protect against diseases caused by bacterial pathogens.
 XX

Example 1; Col 21-22; 28pp; English.

This invention describes a novel recombinant basal body rod protein (I) producible by a transformed host containing an expression vector comprising a nucleic acid selected from; (a) an entire nucleic acid sequence of 1800 base pairs (bp), or the nucleic acid sequence of the flgG gene having 810 bp, both given in the specification; (b) a nucleic acid encoding the amino acid sequence of the flgG protein, a 270 residue sequence, given in the specification; (c) a nucleic acid sequence encoding a functional flgG basal body rod protein of a flagellum of a strain of Campylobacter; or (d) an immunogenic fragment of an flgG protein of (a), (b) or (c), and expression means operatively coupled to the nucleic acid molecule for expression by the host of a basal body rod protein of a flagellum of a strain of Campylobacter. (I) is useful in immunological vaccine compositions for in vivo administration to protect against diseases caused by bacterial pathogens that produce basal body rod proteins. (I) are also useful as immunogens, as antigens in immunoassays, or for procedures for the detection of antibacterial, Campylobacter, basal body rod protein and/or peptide antibodies. The product of the invention has immunostimulatory activity. This sequence encodes the Campylobacter jejuni flgG protein described in the method of the invention

XX SQ Sequence 789 BP; 260 A; 127 C; 179 G; 223 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,29e-10 Length: 789
 Score: 217.50 Matches: 95
 Percent Similarity: 27.45% Conservative: 45
 Best Local Similarity: 18.63% Mismatches: 115
 Query Match: 9.18% Indels: 255
 DB: 3 Gaps: 9

US-10-009-823A-1 (1-502) x AAZ88554 (1-789)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
 Db 1 ATGATGAGATCACTTACTATCTGCTACAGAAATGTTAGCGGACGAAACAAATTCAT 60
 QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
 Db 61 GTTACTTCAATAAATACATCGCCATGTTTAATACAGCAGGTTTAAAGAAAGTCGCGCAA 120
 QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
 Db 121 TTGCTGATCTATGTATCAAGTTATGAAGTATGCGAGAACTTCACTTCAGCTACTACT 180
 QY 61 GlnAlaGlyMetGlyAlaGlnValGly-----SerValArgThrIle 74
 Db 181 CTTTCTCCTTCGGGTATAGAGTGGGTGTGGTGTGCGTCCACAGCGGTAACTAAAGTT 240
 QY 75 PheThrGlnGlyAlaPheGlu---ProGlyAsnSerValThrAspLeuAlaIleGlyGly 93
 Db 241 TTACTGAAGGAATTTAAATCAACAAAGTACTGATGCTTGTATATGGCTATTCAGGT 300
 QY 94 LysGlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsn 112
 Db 301 AATGGGTTTTTCAATACAACTTCTCTGATGGCACTATAGGATATAGAAATGGCAA 360
 QY 113 PheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySer 132
 Db 361 TTTACAAAGATAATGAAGGTATATATTGTAATTCAGATCGTTATAGACTT----- 411
 QY 133 ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe 152
 Db 411 ----- 411
 QY 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal 172
 Db 411 ----- 411
 QY 173 AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeu 192

```
Db 411 ----- 411
QY 193 GluSerTrpLysGlyAsnGlyThrProPheSerThrSerAsnTyrSerTyrAlaGln 212
Db 411 ----- 411
QY 213 ProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGly 232
Db 411 ----- 411
QY 233 AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGlu 252
Db 411 ----- 411
QY 253 AspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThr 272
Db 411 ----- 411
QY 273 PheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThr 292
Db 411 ----- 411
QY 293 LysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAla 312
Db 412 ----- 420
QY 313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332
Db 421 ----- 438
QY 333 AsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352
Db 439 ----- 474
QY 353 ProSerMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 372
Db 475 TCTGTATGCTACCA ----- 489
QY 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392
Db 489 ----- 489
QY 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412
Db 490 ----- 519
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyr 432
Db 520 GTGCGAGCTAGTTCAGTTATAAATCCAGCGGCTTCATCTATGGTGATATCTTAT 579
QY 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 580 CTTGAAACAGGAGCAAGTGTGCACCTGTTCGCGGTATAGCAGCAAGATGGCTTGA 639
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 640 ACAATAGACATCGATTATAGACTAGTAATGTTTCAGCTTGTGAAGAAATGACAGAT 699
QY 473 MetIleIleLeuGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 700 CTTATATCAGGACAAAGAGCTTATAGCGGGTCTTAAGGCCATTACACACAGTGTATGAT 759
QY 493 MetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 760 ATGCTAGGAATTGTAATCAGCTTAAGCGGA 789
```

RESULT 37

ABK74597

ID ABK74597 standard; DNA, 555 BP.

XX ABK74597;

XX ABK74597;

DT 13-AUG-2002 (first entry)

```
XX DE Bacillus licheniformis genomic sequence tag (GST) #1888.
XX KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX OS Bacillus licheniformis.
XX WO200229113-A2.
XX PN 11-APR-2002.
XX PD 05-OCT-2001; 2001WO-US031437.
XX PF 06-OCT-2000; 2000US-00680598.
XX PR 27-MAR-2001; 2001US-0279526P.
XX PA (NOVO ) NOVOZYMES BIOTECH INC.
XX PA (NOVO ) NOVOZYMES AS.
XX PI Berka R, Clausen IG;
XX WIPI; 2002-416684/44.
XX PT Monitoring differential expression of several genes in first Bacillus
XX cell relative to expression of same genes in one or more second Bacillus
XX cells, by using substrate containing Bacillus genomic sequenced tag
XX array.
XX Claim 4; SEQ ID NO 1888; 200pp; English.
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX relative gene expression by an observed hybridisation reporter signal of
XX a spot in the array. The method is useful for measuring the expression of
XX genes in a first Bacillus cell relative to expression of the same genes
XX in one or more second Bacillus cells. The method is useful for monitoring
XX global expression of several genes from a Bacillus cell, discovering new
XX genes, identifying possible functions of unknown open reading frames and
XX monitoring gene copy number variation and stability. Monitoring changes
XX in expression of genes may be used to provide a representation of the way
XX in which Bacillus cells adapt to changes in culture conditions,
XX environmental stress or other physiological provocation. Extensive follow
XX up characterisation is unnecessary, when one spot on an array equals one
XX gene or one open reading frame, since sequence information is available.
XX This sequence represents a genomic sequence tag (GST) used in the method
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 555 BP; 163 A; 145 C; 129 G; 118 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:	3.74e-10	Length:	555
Score:	237.00	Matches:	59
Percent Similarity:	49.70%	Conservative:	25
Best Local Similarity:	34.91%	Mismatches:	59
Query Match:	9.16%	Indels:	26
DB:	6	Gaps:	2

US-10-009-823A-1 (1-502) x ABK74597 (1-555)

QY 360 ThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGln 379

Db 48 ACCATTACAGGAGCGGCTACTTCCGGTCGGATCAGGAGATGAAGTGTACTACACAGA 107

QY 380 AspGly-----TyrProGlnGlyAspLeuValAspVal----- 390

Db 108 TCAGGAAACTTCTACCGATCTGATGAAGTGTGACCTGTCTACTGTTCGGGCTGTTCGTC 167

CC (2) a host cell comprising the vector; (3) an isolated polypeptide
CC encoded by the nucleic acid; (4) an article of manufacture comprising the
CC polypeptide; (5) an antibody having specific binding affinity for the
CC polypeptide; (6) a method for detecting the presence or absence of L.
CC intracellularis in a biological sample; (7) a method of preventing
CC infection by L. intracellularis in an animal; (8) a composition
CC comprising a first oligonucleotide primer and a second oligonucleotide
CC primer, where the first and second primers are each 10 to 50 nucleotides
CC in length, and where in the presence of L. intracellularis nucleic acid,
CC generate an amplification product under standard amplification
CC conditions, but do not generate an amplification product in the presence
CC of nucleic acid from an organism other than L. intracellularis; and (9)
CC an article of manufacture comprising the composition. The nucleic acid
CC sequence has antibacterial activity, and can be used in immunotherapy.
CC The nucleic acid and polypeptides are useful for treating and preventing
CC L. intracellularis infection e.g. proliferative enteropathy. The present
CC sequence represents an L. intracellularis specific nucleic acid molecule,
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 742 BP; 213 A; 149 C; 133 G; 245 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 1.05e-08 Length: 742
Score: 221.00 Matches: 62
Percent Similarity: 48.97% Conservative: 33
Best Local Similarity: 31.96% Mismatches: 87
Query Match: 8.55% Indels: 12
DB: 12 Gaps: 3

US-10-009-823A-1 (1-502) x ADM91627 (1-742)

QY 316 GlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrp 335
DB 682 GGAAGTCGCTTACACCGCTCCATTAATTTTACCAGGGGACTCACAGAAC-----632
QY 336 AlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMet 355
DB 631 ACAGGTAATCTCTAAATATGGCTATTGAGGCAAGGGTTTTCACAGTA-----TTG 578
QY 356 MetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArg 375
DB 577 GTTGTGTATCAATTAATGTATATACAGACGAGCTGCTTTAACTTAATCAGATGGGACT 518
QY 376 ArgTyrSerGlnAspGlyTyrPro-----GlnGlyAspLeuVal 388
DB 517 CTTGTAACAGCAATGATGATACCCCTACACAGAAATTTACAGTTCTCTCTGATGCTAAA 458
QY 389 AspValThrIleThrSerGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValVal 408
DB 457 TCTGTGCTATTCTGAAAGAGGTCGTATTACTGCTTTAGACGCAATGGACAGAAATC 398
QY 409 AspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGly 428
DB 397 GCTGCTGGAGAAATACCTTTATATAGCTTTGTTAATCCAGCGGGACTGGATGCAAAAGA 338
QY 429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThr 448
DB 337 CGTAATCTCTATATTTCCACAGAACATCCGCTGAAGCGGTAGAGGGTGTCCCTCGTGAA 278
QY 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
DB 277 GAAATCTTAGGACATATAGCTCAAGATCTCTGAAATGCTCTAATGTAGAGTTCTGTGAT 218
QY 469 GluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
DB 217 GAAATGGTACAATGATTGTAGGACAAAGAGCATATGAATGAATTCAAAAGCAATCAAA 158
QY 489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
DB 157 ACTTCAGACAAATGCTTCTTAACTGCGATGCCAATCTTAACGT 116

RESULT 39
ABK74555

QY 391 -----ThrIleThr 393
DB 168 TTAACGGCCAATAATGGAAGATCAACATTCGCGAAGACGCCCAAGCTTACGATCGCT 227
QY 394 SerGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
DB 228 CTGACGGGAACCGTTTCTTATGTCGATCAAAACACAGAGACACAGCCGCCGCCAGATC 287
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGlyLysAsnAsnHisTyrSer 433
DB 288 TCGCTTGCACATTCAGCAACATCCGATATCAAAAGCTGGAGACAACTTTACCGC 347
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
DB 348 GAAACATTAGTTCGGAGATCCGCAAGTCGTGTACCTGGGGAAGCGGCTCAGGAAA 407
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
DB 408 ATTCAAACGAGTCGCTTGAATGTGCAACGTCGACCTTTTCAGAAAGATTTTCCGAAATG 467
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
DB 468 ATTATCGGACGCGCGCTTCCAGTCCATGCGAAATCATTAACACCTCTGTGTAAT 527
QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502
DB 528 CTTACAGAACTCGTCAATCTGAAGCGA 554

RESULT 38
ADM91627/c
ID ADM91627 standard; DNA; 742 BP.
XX
AC ADM91627;
XX
DT 15-JUL-2004 (first entry)
XX
DE Lawsonia intracellularis specific nucleotide sequence SEQ ID NO:28.
XX
KW Lawsonia intracellularis; infection; antibacterial; immunotherapy; gene;
KW ds; proliferative enteropathy.
XX
OS Lawsonia intracellularis.
XX
SS WO2004033631-A2.
XX
XX 22-APR-2004.
XX
XX 01-OCT-2003; 2003WO-US031318.
XX
XX 04-OCT-2002; 2002US-0416395P.
XX
XX (MINU) UNIV MINNESOTA.
XX
XX Kapur V, Gebhart CV;
XX
XX WPI; 2004-340902/31.
XX
XX P-PSDB; ADM91689.
XX
XX New nucleic acid that generates an amplification product from L.
XX intracellularis nucleic acid using an appropriate second nucleic acid
XX molecule, useful for treating and preventing L. intracellularis
XX infection.
XX
XX Disclosure; SEQ ID NO 28; 87pp; English.
XX
XX The present invention describes an isolated nucleic acid comprising a
XX nucleic acid molecule of at least 10 nucleotides in length having at
XX least 75% identity to SEQ ID NO:8741 (ADM91733), where any of the
XX molecule that is 10-29 nucleotides in length, under standard
XX amplification conditions, generates an amplification product from
XX Lawsonia intracellularis nucleic acid using an appropriate second nucleic
XX acid molecule. Also described: (1) a vector comprising the nucleic acid;

QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleYsSerGlnGlnAen 333
DB 1039 TTTTACCTTCGCGGA---GAACCGGAGGAGCATTTCTTCAATTTACCGGAGGAGAG--- 1092
QY 334 MetTrpAlaProAlaSerAlaAlaIleGlyThrAspIleGlyIleValPro 353
DB 1093 ---GCTGAGCTGCAAGGGAGCGCC---GCAAAATCAAA 1128
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg-----AsnGlySer 371
DB 1129 GTAGCCGATGATCATGATTCAAAGGGGATAAAGTCGCCGATCGCTGATGTGAA 1188
QY 372 SerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspValValThr 391
DB 1189 GCGAGC-----GACAACGCCAACCGCAATCTCGTAAATGCTTC 1230
QY 392 IleThrSerGluGlyIleValGlnGlyTyrSerAsnSerGlnValValAspPheTyr 411
DB 1231 ACCAAG-----AAATTCAAATCGGTGACAAACGACCGCTTTAGACTATTAC 1281
QY 412 AsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHis 431
DB 1282 -----GAGGCGATCATCGGTGAATGGCGTAAAGCTCAGGAAACAAACAGA 1329
QY 432 TyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyr 451
DB 1330 CTGCCCAAAATACAGAAACG-----CTTGTAACACTGCTGATTC 1371
QY 452 GlyIleValSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetVal 471
DB 1372 AACAGACAGTCTGTT-----AGCGCCGTATCATTAGACGAAGAATGTCG 1416
QY 472 AsnMetIleIleGlnArgGlyPheGlnMetAsnSerIleYsSerValThrAlaAsp 491
DB 1417 AACATGATCAATTCAGATGCTTATACCGCGCTCGGAGATGATTACGCTGCGAGGAC 1476
QY 492 ThrMetLeuGlnIleAlaLeu 498
DB 1477 GAAGTCTCGATAGATTAT 1497
RESULT 40
ID ADG32119 standard; DNA; 1432 BP.
AC AC
XX ADG32119;
DT 26-FEB-2004 (first entry)
TX DNA encoding a mutant B_licheniformis secreted polypeptide SeqID 89.
DE mutant; host cell; production yield; shelf life; product stability;
KW purity; secreted; gene; ds.
XX Synthetic.
OS Bacillus licheniformis.
XX WO20030393453-A2.
XX 13-NOV-2003.
XX 25-MAR-2003; 2003WO-DK000198.
XX 10-APR-2002; 2002DK-00000534.
XX (NOVO) NOVOZYMES AS.
XX Andersen JT, Jorgensen ST, Rasmussen MD, Olsen PB, Clausen IG;
XX WPI; 2004-053045/05.
XX P-PSDB; ADG32120.
XX New mutant Bacillus licheniformis host cell secreting 5 % less of one or

PT more secreted polypeptides than the parent host cell, useful for
PT producing a product of interest e.g. polypeptides, amino acids or
XX carbohydrates.
PS Disclosure; SEQ ID NO 89; 422pp; English.
XX This invention relates to a novel Bacillus licheniformis (B.
CC licheniformis) mutant host cell derived from a parent B. licheniformis
CC host cell that is mutated in genes encoding secreted polypeptides.
CC Specifically, it refers to the generation of an improved Bacillus host
CC that reduces the need for product purification caused by contaminant
CC secreted native polypeptides in the culture medium. Accordingly, the
CC present invention describes reducing the expression of these native
CC proteins (e.g. proteolytic enzymes, nutrient uptake factors and signal
CC molecules), which in turn makes it easier to purify the heterologous
CC product of interest and therefore improving the production process.
CC Further benefits of a mutated host cell include an increase in total
CC production yield and a longevity of shelf life attributable to improved
CC product stability and purity. This polynucleotide is a DNA sequence
CC encoding a mutant B. licheniformis secreted polypeptide of the invention.
SQ Sequence 1432 BP; 477 A; 287 C; 335 G; 333 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 9.82e-07 Length: 1432
Score: 201.00 Matches: 40
Percent Similarity: 66.25% Conservative: 13
Best Local Similarity: 50.00% Mismatches: 27
Query Match: 7.77% Indels: 0
DB: 12 Gaps: 0
US-10-009-823A-1 (1-502) x ADG32119 (1-1432)
QY 423 GlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGlu 442
DB 4 GGTATTATCAAAAGCTGGAGACAACTTTACCGCGAAACATTGAGTTCGGAGATCGCAA 63
QY 443 PheGlyLeuProGlyThrSerAsnTyrGlyIleValSerValAsnGlnLeuGluThrSer 462
DB 64 GTCGTTGTACTTGGGAAAGCGGCTCAGGAAATAATCAACGAGTGGCGCTTGAATGTCG 123
QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482
DB 124 AACGTCGACCTTCAGAGAGATTTCCGAATGATTTCCGCGAGCGGGCTTCAGTCC 183
QY 483 AsnSerIleValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
DB 184 AATCGMAATCATTAACACTCTGATGAATTTCTCAGGAACCTGCTCAATCTGAAGCGA 243

Search completed: November 1, 2004, 05:38:16
Job time : 912 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:35:12 ; Search time 195 Seconds

(without alignments)
531.115 Million cell updates/sec

Title: US-10-009-916A-1

Perfect score: 950

Sequence: 1 MKIKLFFVTSIVTISLLTSI.....DKPLPLGGGARIACGVIPN 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	476.5	50.2	178	2 Q7UCE0	Q7UCE0 shigella fl
2	476.5	50.2	190	2 Q83RB4	Q83RB4 shigella fl
3	474.5	49.9	173	1 SODC ECOLI	P53635 escherichia
4	473.5	49.8	190	2 Q8FH80	Q8FH80 escherichia
5	452	47.6	173	2 Q704S6	Q704S6 salmonella
6	452	47.6	173	2 CAF06532	CAF06532 salmonella
7	451	47.5	173	1 SODC_SALTY	Q88901 salmonella
8	449	47.3	173	2 Q8ZG66	Q8ZG66 salmonella
9	435.5	45.8	172	2 Q887F8	Q887F8 pseudomonas
10	425	44.7	177	1 SODI_SALTY	P53636 salmonella
11	425	44.7	177	2 Q704S7	Q704S7 salmonella
12	425	44.7	177	2 CAF06531	CAF06531 salmonella
13	420	44.2	174	1 SODC_BRUME	P58845 brucella me
14	417.5	43.9	174	2 Q6D531	Q6D531 erwinia car
15	414	43.6	171	2 Q9KRQ3	Q9KRQ3 vibrio chol
16	414	43.6	175	2 Q7AFX5	Q7AFX5 escherichia
17	414	43.6	274	2 Q8X9P0	Q8X9P0 escherichia
18	413.5	43.5	154	1 SODC_BRUAB	P15453 brucella ab
19	413.5	43.5	174	2 Q9MC02	Q9MC02 phage fels-
20	413.5	43.5	174	2 Q8ZQF7	Q8ZQF7 salmonella
21	413	43.5	175	2 Q7AAL1	Q7AAL1 escherichia
22	413	43.5	328	2 Q8XCB6	Q8XCB6 escherichia
23	411	43.3	170	2 Q7MFM9	Q7MFM9 vibrio vuln
24	410	43.2	171	2 Q87G06	Q87G06 vibrio para
25	407	42.8	170	2 Q8D454	Q8D454 vibrio vuln
26	404	42.5	195	2 Q8D1A0	Q8D1A0 yersinia pe
27	404	42.5	201	2 Q8ZBN3	Q8ZBN3 yersinia pe
28	404	42.5	201	2 AAS60586	AAS60586 yersinia
29	399	42.0	170	2 Q7VWV9	Q7VWV9 bordetella
30	391	41.2	170	2 Q7WH53	Q7WH53 bordetella
31	391	41.2	173	1 SODC_PHOLE	P00446 photobacter

ALIGNMENTS

RESULT 1

Q7UCE0 PRELIMINARY; PRT; 178 AA.

ID Q7UCE0
AC Q7UCE0;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Superoxide dismutase (Cu-Zn)
GN Name=sodC; OrderedLocusNames=S1805;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24577;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 24577";
RL Infect. Immun. 71:2775-2786(2003).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2O)(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL; AB016983; AAPI7140.1; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; SodCu; 1_CU_ZN.
DR PRODOM; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 178 AA; 18167 MW; DC43B64712B8FA65 CRC64;

Query Match 50.2%; Score 476.5; DB 2; Length 178;

Best Local Similarity 54.1%; Pred. No. 6.5e-35;

Matches 92; Conservative 24; Mismatches 51; Indels 3; Gaps 2;

QY 10 SIVTISLLTSITSVVLACSVTSEVHMDDNGIKQISGIVTFTDQKGLQIKTDKGLPAG 69

Db 10 SLAILALV--VATGAQAASEKVENMLVTSQGVSGISGIVTITETDKGLRSPDLKALPPG 67

QY 70 BGFHIIHGSCGPAEHDGHLTAGLQACHYDPPKTKHGKPLNGHKGLPLRLVAVKADG 129

Db 68 BGFHIIHGSCGPAEHDGHLTAGLQACHYDPPKTKHGKPLNGHKGLPLRLVAVKADG 127

QY 130 IAKETLLAPRL-TVKEIKGRVTMIHAGDNYSDKPLPLGGGGARIACGVI 178

RA Benov L.T., Chang L.Y., Day B., Fridovich I.;
 RT "Copper, zinc superoxide dismutase in Escherichia coli: periplasmic
 RT localization.";
 RL Arch. Biochem. Biophys. 319:508-511(1995).
 RN [9]
 RP CHARACTERIZATION.
 RX MEDLINE=96063713; PubMed=7589534;
 RA Battistoni A., Rotilio G.;
 RA "Isolation of an active and heat-stable monomeric form of Cu,Zn
 RT superoxide dismutase from the periplasmic space of Escherichia coli.";
 RL FEBS Lett. 374:199-202(1995).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 20-173.
 RX MEDLINE=98070606; PubMed=9405149;
 RA Pesce A., Capasso C., Battistoni A., Folcarelli S., Rotilio G.,
 RA Desideri A., Bolognesi M.;
 RT "Unique structural features of the monomeric Cu,Zn superoxide
 RT dismutase from Escherichia coli, revealed by X-ray crystallography.";
 RL J. Mol. Biol. 274:408-420(1997).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems. This enzyme is
 CC highly thermostable.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U51242; AAC03729.1; -;
 DR EMBL; A3000259; AAC74718.1; -;
 DR EMBL; A3005387; AAG56635.1; -;
 DR EMBL; A3002558; BAB35778.1; -;
 DR EMBL; X97766; CAA66363.1; -;
 DR PIR; C90923; C90923.
 DR PIR; G85771; G85771.
 DR PIR; JC6004; JC6004.
 DR PDB; 1ESO; X-ray; @=20-173.
 DR EcoBASE; EB3195; -;
 DR Scogen; EGI3419; sodC.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; Sod Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR ProSITE; PS00087; SOD_CU_ZN_1; FALSE_NEG.
 DR ProSITE; PS00332; SOD_CU_ZN_2; 1.
 KW 3D-structure: Antioxidant; Complete proteome; Copper;
 KW Direct protein sequencing; Metal-binding; Oxidoreductase; Periplasmic;
 KW Signal; Zinc.
 FT SIGNAL 1 19
 FT CHAIN 20 173 Superoxide dismutase [Cu-Zn].
 FT METAL 67 67 Copper.
 FT METAL 69 69 Copper.
 FT METAL 92 92 Copper and zinc.
 FT METAL 101 101 Zinc.
 FT METAL 109 109 Zinc.
 FT METAL 112 112 Zinc.
 FT METAL 147 147 Copper.
 FT DISULFID 74 169
 FT STRAND 21 30
 FT TURN 31 32
 FT STRAND 33 46
 FT TURN 47 48
 FT STRAND 49 56
 FT STRAND 62 69
 FT STRAND 66 69
 FT STRAND 77 77

FT STRAND 84 84
 FT TURN 85 86
 FT HELIX 87 89
 FT STRAND 92 92
 FT TURN 95 96
 FT TURN 104 105
 FT TURN 110 111
 FT STRAND 112 112
 FT STRAND 116 118
 FT TURN 120 121
 FT STRAND 124 124
 FT STRAND 128 130
 FT TURN 131 132
 FT HELIX 136 139
 FT TURN 140 141
 FT STRAND 142 147
 FT HELIX 159 162
 FT STRAND 166 173
 SQ SEQUENCE 173 AA; 17681 MW; 9A0CB65F03AAB197 CRC64;
 Query Match 49.9%; Score 474.5; DB 1; Length 173;
 Best Local Similarity 54.1%; Pred. No. 9.5e-35;
 Matches 92; Conservative 24; Mismatches 51; Indels 3; Gaps 2;
 QY 10 SIVTISLTSTSVVLACSVTSEVHMIDMDNGIKQSIGTFTDTDKGLQIKTLKGLPAG 69
 DB 5 SLATLALV--VATGAQAASEKVENLVTSQGVSGISGVITETDKGLFSPDLKALPPG 62
 QY 70 EHGPHIEGSGCGPAEHGHLTAGLQAHGHYDDPKTKHKGPLNGHKGDLPLRVVVKADG 129
 DB 63 EHGPHIEGSGCGPATKDGKASAEAGHLDPQNTKGHEGPEGAGHLGDLPALVWVNDG 122
 QY 130 IAKETTLAPRL--TVKEIKGRTVMHAGGDNVSDKPLPLGGGGARIACGVI 178
 DB 123 KATDAVIAPRLKSLDEIKDALMVHVGDNMSDQPKPLGGGGERYACGVI 172
 RESULT 4
 Q8FH80 PRELIMINARY; PRT; 190 AA.
 ID Q8FH80
 AC Q8FH80;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Superoxide dismutase (EC 1.15.1.1).
 GN Name=sodC; OrderedLocusNames=c2038;
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;
 EX MEDLINE=23388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AE016761; AAN80498.1; -;
 DR HSSP; P53635; IESG.
 DR GO; GO:0004785; P:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; P:metal ion binding; IEA.
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.

```

DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; SOD_Cu; 1..ZN; 1.
DR ProDom: PD000469; SOD_CU_ZN; 1.
DR PROSITE: PS00332; SOD_CU_ZN; 2; 1.
KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 190 AA; 19617 MW; 6217P2DAAE1A8E23 CRC64;

Query Match 49.8%; Score 473.5; DB 2; Length 190;
Best Local Similarity 53.5%; Pred. No. 1.3e-34;
Matches 91; Conservative 25; Mismatches 51; Indels 3; Gaps 2;

Qy 10 SIVTSLTTSITSVLACSVTSEVMHDDNGIKQSIGTFTTDTKGLQIKTKLPGAG 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 22 SLAILLV--VATQAQAASEKVENMLVTSQVGSGISGVITITDKGLESPDLKALPPG 79

Qy 70 EHGHEHIGGSCGPAEDHGLTAGLQAHGHYDPDKTGKHEGPGPLNGHKGDLPLVVKAG 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 EHGHEHIGGSCGPAEDHGLTAGLQAHGHYDPDKTGKHEGPGAGHLGDLPLVVDNG 139

Qy 130 IAKETLLAPRL-TVKEIKGRTVMIHAGDNYSDKPLPLGGGGARIACGVI 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 KATDAVAPRLKSLDEVKDALMHVHGDNMSDQPKPLGGGGERYACGVI 189

RESULT 5
ID Q704S6 PRELIMINARY; PRT; 173 AA.
AC Q704S6;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Superoxide dismutase precursor (EC 1.15.1.1).
GN Name=sodC2;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=119912;
RN [1]
RP SEQUENCE FROM N.A.
RA Amendola S., Ajello M., Pasquali P., Kroll J.S., Langford P.R.,
RA Rotilio G., Valenti P., Battistoni A.;
RT "Differential contribution of sodC1 and sodC2 to intracellular
RT survival and pathogenicity of Salmonella enterica subsp. enterica
RT serovar Choleraesuis.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ620904; CAF06532.1; -.
KW Oxidoreductase; Signal.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 173 superoxide dismutase [Cu-Zn].
SQ SEQUENCE 173 AA; 17721 MW; 511A19B8057F1255 CRC64;

Query Match 47.6%; Score 452; DB 2; Length 173;
Best Local Similarity 52.3%; Pred. No. 1e-32;
Matches 91; Conservative 22; Mismatches 49; Indels 12; Gaps 3;

Qy 12 VTISLTSITSVLAC-----SVTSEVMHDDNGIKQSIGTFTTDTKGLQIKTKLKG 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 LSLAWT-----LLACGAQAASEKVENMLVTAQGVGSGITVVIDETGGKFTPHKA 58

Qy 66 LPAGEHGFHIHGGSCGPAEDHGLTAGLQAHGHYDPDKTGKHEGPGPLNGHKGDLPLVV 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 LPPCEHGFHIHANGSCQPAIKDQAAVAAGHLDLPQNTGKHEGPGQGHLDPLVV 118

Qy 126 KADGIAKETLLAPRL-TVKEIKGRTVMIHAGDNYSDKPLPLGGGGARIACGVI 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 NNDGIATEPVTAPRLKSLDEVKDALMHVHGDNMSDQPKPLGGGGERYACGVI 172

RESULT 7
SODC_SALTY
ID SODC_SALTY STANDARD; PRT; 173 AA.
AC O68901;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 2 precursor (EC 1.15.1.1) (sodCII).
GN Name=sodC; Synonyms=sodC2; OrderedLocusNames=STM1440;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 14028S;
RA MEDLINE=99307439; PubMed=10377444;
RA Fang F.C., DeGroot M.A., Foster J.W., Baunler A.J., Ochsen U.,
RA Testerman T., Bearson S., Giard J.C., Xu Y., Campbell G., Laessig T.;
RT "Virulent Salmonella typhimurium has two periplasmic Cu, Zn-superoxide
RT dismutases.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:7502-7507(1999).
RN [2]
RP SEQUENCE FROM N.A.

```

```

Db 119 NNDGIATEPVTAPRLKSLDEVKDALMHVHGDNMSDQPKPLGGGGERYACGVI 172

RESULT 6
CAF06532 PRELIMINARY; PRT; 173 AA.
ID CAF06532;
AC CAF06532;
DT 10-MAY-2004 (TRENBLrel. 27, Created)
DT 10-MAY-2004 (TRENBLrel. 27, Last sequence update)
DT 10-MAY-2004 (TRENBLrel. 27, Last annotation update)
DE Superoxide dismutase precursor (EC 1.15.1.1).
GN SODC2.
OS Salmonella enterica subsp. enterica serovar Choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella; Salmonella enterica.
OX NCBI_TaxID=119912;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A50;
RA Amendola S., Ajello M., Pasquali P., Kroll J.S., Langford P.R.,
RA Rotilio G., Valenti P., Battistoni A.;
RT "Differential contribution of sodC1 and sodC2 to intracellular
RT survival and pathogenicity of Salmonella enterica subsp. enterica
RT serovar Choleraesuis.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ620904; CAF06532.1; -.
KW Oxidoreductase; Signal.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 173 superoxide dismutase [Cu-Zn].
SQ SEQUENCE 173 AA; 17721 MW; 511A19B8057F1255 CRC64;

Query Match 47.6%; Score 452; DB 2; Length 173;
Best Local Similarity 52.3%; Pred. No. 1e-32;
Matches 91; Conservative 22; Mismatches 49; Indels 12; Gaps 3;

Qy 12 VTISLTSITSVLAC-----SVTSEVMHDDNGIKQSIGTFTTDTKGLQIKTKLKG 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 LSLAWT-----LLACGAQAASEKVENMLVTAQGVGSGITVVIDETGGKFTPHKA 58

Qy 66 LPAGEHGFHIHGGSCGPAEDHGLTAGLQAHGHYDPDKTGKHEGPGPLNGHKGDLPLVV 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 LPPCEHGFHIHANGSCQPAIKDQAAVAAGHLDLPQNTGKHEGPGQGHLDPLVV 118

Qy 126 KADGIAKETLLAPRL-TVKEIKGRTVMIHAGDNYSDKPLPLGGGGARIACGVI 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 NNDGIATEPVTAPRLKSLDEVKDALMHVHGDNMSDQPKPLGGGGERYACGVI 172

RESULT 7
SODC_SALTY
ID SODC_SALTY STANDARD; PRT; 173 AA.
AC O68901;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 2 precursor (EC 1.15.1.1) (sodCII).
GN Name=sodC; Synonyms=sodC2; OrderedLocusNames=STM1440;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 14028S;
RA MEDLINE=99307439; PubMed=10377444;
RA Fang F.C., DeGroot M.A., Foster J.W., Baunler A.J., Ochsen U.,
RA Testerman T., Bearson S., Giard J.C., Xu Y., Campbell G., Laessig T.;
RT "Virulent Salmonella typhimurium has two periplasmic Cu, Zn-superoxide
RT dismutases.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:7502-7507(1999).
RN [2]
RP SEQUENCE FROM N.A.

```

```

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DT
DE Copper-zinc superoxide dismutase.
DE
DE Name=sodC; OrderedLocusNames=STY1692, t1308;
DE
DE Salmonella typhi.
DE
DE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
DE
DE Enterobacteriaceae; Salmonella.
DE
DE NCBI_TaxID=601;
DE
DE (1)
DE
DE SEQUENCE FROM N.A.
DE
DE RN STRAIN=CT18;
DE
DE RC STRAIN=CT18;
DE
DE RX MEDLINE=2534947; PubMed=11677608; DOI=10.1038/35101607;
DE
DE RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
DE
DE RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
DE
DE RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
DE
DE RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
DE
DE RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
DE
DE RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
DE
DE RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
DE
DE RA Whitehead S., Barrrell B.G.;
DE
DE RT "Complete genome sequence of a multiple drug resistant Salmonella
DE
DE RT enterica serovar Typhi CT18.";
DE
DE RL Nature 413:848-852(2001).
DE
DE RN (2)
DE
DE RN SEQUENCE FROM N.A.
DE
DE RP STRAIN=Ty2 / ATCC 700931;
DE
DE RC MEDLINE=22531367; PubMed=12644504;
DE
DE RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
DE
DE RA Burland V., Kodyonanni V., Schwartz D.C., Blattner F.R.;
DE
DE RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
DE
DE RT and CT18.";
DE
DE RL J. Bacteriol. 185:2330-2337(2003).
DE
DE CC -!- FUNCTION: Destroys radicals which are normally produced within the
DE
DE CC cells and which are toxic to biological systems (By similarity).
DE
DE CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
DE
DE CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
DE
DE CC similarity).
DE
DE CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DE
DE DR EMBL; AL627271; CAD01927.1; -.
DE
DE DR EMBL; AE016838; AA068958.1; -.
DE
DE DR HSP; P53635; IES0.
DE
DE DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DE
DE DR GO; GO:0046872; F:metal ion binding; IEA.
DE
DE DR GO; GO:0006801; P:superoxide metabolism; IEA.
DE
DE DR InterPro; IPR001424; SOD_CU_ZN.
DE
DE DR Pfam; PF0080; Sod_Cu; 1.
DE
DE DR ProDom; PD000469; SOD_CU_ZN; 1.
DE
DE DR PROSITE; PS00332; SOD_CU_ZN 2; 1.
DE
DE KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
DE
DE SQ SEQUENCE 173 AA; 17335 MW; 00A5A8B56F25B4EF CRC64;
DE
DE
DE Query Match 47.3%; Score 449; DB 2; Length 173;
DE
DE Best Local Similarity 52.3%; Pred. No. 1.9e-32;
DE
DE Matches 91; Conservative 22; Mismatches 49; Indels 12; Gaps 3
DE
DE
DE QY 12 VTISLLTISITSVLAC-----SVTSVHMIDNDNGIKQSIGTFTDTDKGLQIKTLDKG 65
DE
DE DB 4 LSLAVMT-----LIACAGQAASEKVENMLVITAGVGQSIGTIVVDETEGGLKFTPLKA 58
DE
DE
DE QY 66 LPAGEHGFIHEGSCQPAEDGHGLTAGLOAHGHYDPDKTKHGEPGLNGHKGDLPLRLV 125
DE
DE DB 59 LPPEHGFIHAKGSCQPAIKDGKAVAAEAGGHLDPQNTGKHGEPGQCHGLDPLVLV 118
DE
DE
DE QY 126 KADGTAKETLLAPRL-TVKETKGTFTVIHAGDNYSDKPLPLGGGGARITACGVI 178
DE
DE DB 119 NNDGATPEPTVAPRLKSLDEVKALKMIRHVGGDNMSDQPKPLGGGTRITACGVI 172
DE
DE
DE RESULT 9
DE
DE Q887F8 PRELIMINARY; PRT; 172 AA.
DE
DE ID Q887F8;
DE
DE AC Q887F8;
DE
DE DT 01-JUN-2003 (TrEMBLrel. 24, Created)

```

DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Superoxide dismutase, Cu-Zn.
 GN Name=sodC; OrderedLocusNames=SPPT01338;
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OC NCBI_TaxID=333;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RA MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
 RX Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
 RA Winn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
 RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
 RA Nelson W.C., DavidSEN T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
 RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
 RA Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
 RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
 RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
 RA Bender C.L., White O., Fraser C.M., Collier A.;
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen
 Pseudomonas syringae pv. tomato DC3000";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186 (2003).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AF016860; AA054860.1; -.
 DR HSSP; P53636; 1EQW.
 DR TIGR; PSPT01338; -.
 DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR InterPro; IPR010916; TONE_Box_N.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
 DR PROSITE; PS00332; SOD_CU_ZN_2; UNKNOWN_1.
 DR PROSITE; PS00430; TONE_DEPENDENT_REC_1; UNKNOWN_1.
 KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
 SQ SEQUENCE 172 AA; 17465 MW; C8298C408590AA72 CRC64;
 Query Match 45.8%; Score 435.5; DB 2; Length 172;
 Best Local Similarity 51.8%; Pred. No. 3.1e-31;
 Matches 86; Conservative
 Qy 14 ISLLTSTSVLACSVTSSEVHMDNGIKOSIGTVFTDTDKGLQIKTDLKGLPAGEHGF 73
 Db 7 LGLLAPSLVQAAASLDPVNLVSADGAPKLIGSVTSEYGLLTPKLNGLPAGHGF 66
 Qy 74 HIEHGSGCPAEHDGHLTAGLQAHGHYDPTKCKEGLNGHKGDLPLVWKADGIAKE 133
 Db 67 HVHNGSCGAGTKDGVKVAALAGGHDFPAKTGKHLGPVARGHLGDLPALYVTDGTADY 126
 Qy 134 TLIALPLR-TVKEIKGRVTWTHAGDNYSDKPLPLGGGGARIACGVI 178
 Db 127 PVILAPRLKSLSEIKGHVAMVHAGDNDHAMPKPLGGGGDRAACGVI 172
 RESULT 10
 SOD1_SALTY
 ID -SOD1_SALTY STANDARD; PRT; 177 AA.
 AC P53636; C33803; O50545;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Superoxide dismutase [Cu-Zn] 1 precursor (EC 1.15.1.1) (sodC1).
 GN Name=sodC1; Synonyms=sodC; OrderedLocusNames=STM1044;
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OC NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ST4/74;
 RX MEDLINE=98025474; PubMed=9379906;
 RA Farrant J.B., Sansone A., Canvin J.R., Pallen M.J., Langford P.R.,
 RA Wallis T.S., Dougan G., Kroll J.S.;
 RT "Bacterial copper- and zinc-cofactored superoxide dismutase
 contributes to the pathogenesis of systemic salmonellosis";
 RL Mol. Microbiol. 25:785-796 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98054349; PubMed=9391141;
 RA De Groote M.A.; Ochsen U.A.; Shloh M.U.; Nathan C.; McCord J.M.,
 RA Dinuer M.C., Libby S.J., Vazquez-Torres A., Xu Y., Fang F.C.;
 RT "Periplasmic superoxide dismutase protects Salmonella from products of
 phagocyte NADPH-oxidase and nitric oxide synthase";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:13997-14001 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2";
 RL Nature 413:852-856 (2001).
 RN [4]
 RP SEQUENCE OF 77-163 FROM N.A.
 RC STRAIN=ST4/74;
 RX MEDLINE=97023146; PubMed=8869506;
 RA Canvin J., Langford P.R., Wilks K.E., Kroll J.S.;
 RT "Identification of sodC encoding periplasmic [Cu,Zn]-superoxide
 dismutase in Salmonella";
 RL FEMS Microbiol. Lett. 136:215-220 (1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=20428907; PubMed=10970746; DOI=10.1006/jmbi.2000.4074;
 RA Pesce A., Battistoni A., Stroppolo M.E., Polizio F., Nardini M.,
 RA Kroll J.S., Langford P.R., O'Neill P., Sette M., Desideri A.,
 RA Bolognesi M.;
 RT "Functional and crystallographic characterization of Salmonella
 typhimurium Cu,Zn superoxide dismutase coded by the sodC1 virulence
 gene";
 RL J. Mol. Biol. 302:455-478 (2000).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- MUSCULARITY: Encoded by a cryptic bacteriophage.
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 EMBL; Y13121; CAA73588.1; -.
 EMBL; AF007380; AAB62385.1; -.
 EMBL; AE008743; AAL19978.1; -.
 EMBL; X94327; CAA63988.1; -.
 PDB; 1EQW; X-ray; A/B/C/D=22-177.
 StyGene; SG10705; sodC1.

DR	EMBL; AE009694; AAL53823.1; -.
DR	EMBL; AE014566; AAN33888.1; ALT_INIT.
DR	PIR; AD3582; AD3582.
DR	HSP; P63636; LEQW.
DR	TIGR; BRA0703; -.
DR	InterPro; IPR001424; SOD_CU_ZN.
DR	Pfam; PF00080; Sod Cu; 1.
DR	ProDom; PD000469; SOD_CU_ZN; 1.
DR	PROSITE; PS00087; SOD_CU_ZN_1; 1.
DR	PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW	Antioxidant; Complete proteome; Copper; Metal-binding; Oxidoreductase;
KW	Periplasmic; Signal; Zinc.
FT	SIGNAL 1 20 Potential.
FT	CHAIN 21 174 Superoxide dismutase [Cu-Zn].
FT	METAL 68 68 Copper (By similarity).
FT	METAL 70 70 Copper (By similarity).
FT	METAL 93 93 Copper and zinc (By similarity).
FT	METAL 102 102 Zinc (By similarity).
FT	METAL 110 110 Zinc (By similarity).
FT	METAL 113 113 Zinc (By similarity).
FT	METAL 148 148 Copper (By similarity).
FT	DISULFID 75 170 By similarity.
SEQ	SEQUENCE 174 AA; 18262 MW; 08B355A9A6F7F67A CRC64;
Query Match 44.2%; Score 420; DB 1; Length 174;	
Best Local Similarity 49.4%; Pred. No. 7.8e-58;	
Matches 84; Conservative 22; Mismatches 50; Indels 6; Gaps 2;	
QY	15 SLTTSITSVLAC-----SVTSEVHMDDNGIKQSIGVTFTDTDKGLQIKTDLKGLPAG 69
Db	4 SLFIASIVLMAPFAPEASTTVKMYEALTGPGEKVGIWISAEFGGLHFKVMKEKLTPG 63
QY	70 ENGPHIHEGGSCGAHHDCHLTAGLOAHGHYDPDXTKGHEGPLGNCHKGDLPRLVVKADG 129
Db	64 YGFHVHENPSCAPEKQKGVPAALAAAGHYDPGNTHHHLGPEGDHGMDLPRLSANADG 123
QY	130 IAKETLLAPRL-TVKEIKGRTWIHAGDNYSKDPLPLGGGGARACGVI 178
Db	124 KYSEIVVAPHLKLAIEIKQSLMVHGVDNSDKPEPLGGGGARACGVI 173
RESULT 14	
Q6D531	PRELIMINARY; PRT; 174 AA.
ID Q6D531	
AC Q6D531	
DT 01-OCT-2004	(TREMBLrel. 28, Created)
DT 01-OCT-2004	(TREMBLrel. 28, Last sequence update)
DT 01-OCT-2004	(TREMBLrel. 28, Last annotation update)
DE	Copper-zinc superoxide dismutase.
GN	OppNames=ECA2211;
OS	Erwinia carotovora subsp. atroseptica SCRI11043.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Pectobacterium.
OX	NCBI_TaxID=218491;
RN	[1]
RC	SEQUENCE FROM N.A.
RP	STRAIN=SCRI11043;
RA	Bell K.S., Sebahian M., Pritchard L., Holden M., Hyman L.J.,
RA	Holeva M.C., Thomson N.R., Bentley S.D., Churcher C., Mungall K.,
RA	Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,
RA	Frazer A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,
RA	Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,
RA	Salmond G.P.C., Birch P.R.J., Barrrell B.G., Parkhill J., Toth I.K.;
RL	Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; BX950851; CAG75112.1; -.
SEQ	SEQUENCE 174 AA; 17907 MW; 3655E8E2C8355949 CRC64;
Query Match 43.9%; Score 417.5; DB 2; Length 174;	
Best Local Similarity 47.8%; Pred. No. 1.3e-29;	
Matches 86; Conservative 22; Mismatches 63; Indels 9; Gaps 3;	
QY	1 MKILFFVTSTVTLSTITSVVLGASVTSEVHMDDNGIKQSIGVTFTDTDKGLQIK 60

```
Db 1 MKLKALILATGVF-----FSCSTLAASVTTLNEALPTGAGSSIGDVSITETPYGLLFT 53
QY 61 TDLKGLPAGEHGFPHIHEGSCGPAEDHGLTAGLOAHGHYDPDKTGKHEGPGLN-GHKGD 119
Db 54 PNLKGLGELHGHVHENSCEPAEQDGKSVFALAAGGHFDPKTKGHLGPDYDQGHVD 113
QY 120 LPRLVVKADGIKAKETLLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIACGVI 178
Db 114 LPLGLVANADGTSTYPIVLPRIKSLFVQKHALMVHVGDNADYDMEAKLGGGGARMACGVI 173

RESULT 15
ID Q9KRQ3 PRELIMINARY; PRT; 171 AA.
AC Q9KRQ3;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Superoxide dismutase, Cu-Zn.
GN OrderedLocusNames=VC1583;
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Kickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I.,
RA Sellers P., McDonald L.A., Uitterback L.R., Fleischmann R.D.,
RA Nierman W.C., White O., Salzberg S.B., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae";
RL Nature 406:477-483 (2000).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL; AB004235; AAF94737.1; -.
DR PIR; A82183; A82183.
DR HSSP; P00446; 1YAI.
DR TIGR; VC1583; -.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; SOD_Cu; 1.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
KW SEQUENCE 171 AA; 17697 MW; 7483250CEA266C79 CRC64;

Query Match 43.6%; Score 414; DB 2; Length 171;
Best Local Similarity 48.6%; Pred. No. 2.7e-29;
Matches 85; Conservative 19; Mismatches 59; Indels 12; Gaps 2;

QY 5 LFFVTSIVTISLTISITSVLACSVTSEVHMIDNGIKQSIGTFTDTDKGLQIKTDLK 64
Db 6 LFAVAFSSVLAEMIVMT-----DLSSGQSVGTVTVDSEYGTFTPQLT 54
QY 65 GLPAGEHGFHHEGSCGPAEDHGLTAGLOAHGHYDPDKTGKHEGPGLN-GHKGDLPRL 123
Db 55 GLPAGLHGFHVGANGSCETSSKDGKTVLGGAGGHYDPQNTGKRGYPMTNDHGLDLPAL 114
QY 124 VVKADGIKAKETLLAPRLTVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIACGVI 178
Db 115 YVDAQGANQPVILASRFMAEVKGMALMVHAGGDNHSDHPNPLGGGGARIVCGVI 169
```

```
RESULT 16
Q7AFX5 PRELIMINARY; PRT; 175 AA.
ID Q7AFX5;
AC Q7AFX5;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Putative copper/zinc-superoxide dismutase.
GN OrderedLocusNames=ECs1120;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL; AP002554; BAB34543.1; -.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; SOD_Cu; 1.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
DR Copper; Metal-binding; Oxidoreductase; Zinc.
KW SEQUENCE 175 AA; 18259 MW; 00D8A6F5B7B94792 CRC64;

Query Match 43.6%; Score 414; DB 2; Length 175;
Best Local Similarity 47.8%; Pred. No. 2.7e-29;
Matches 88; Conservative 27; Mismatches 53; Indels 15; Gaps 5;

QY 1 MKIKLFFVTSIVTISLTISITSVLACSVTSE-----VHMIDNGIKQSIGTFTDTDKG 56
Db 1 MKCKI-----IAAIAMLTA-----ASCGYAAEQEYPMNLVSADGKEVSGIKTIQETPYG 50
QY 57 LQIKTDLKLPAGEHGFHHEGSCGPAEDHGLTAGLOAHGHYDPDKTGKHEGPGLN-G 115
Db 51 LLFTPALHSLSEGIHGFHVEKGNCPALCKGKPVAAALSAGHFPDKTKGHLGFWSPDG 110
QY 116 HKGDLPLRVVKADGIKAKETLLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIA 174
Db 111 HPGDLPALFVTHDGKANYVPLAPRLNSLKEIKGRSLMLHAGGDNHHDHPEPLGGGGARMA 170
QY 175 CGVI 178
Db 171 CGII 174

RESULT 17
Q8X9P0 PRELIMINARY; PRT; 274 AA.
ID Q8X9P0;
AC Q8X9P0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Putative copper-zinc superoxide dismutase encoded within prophage CP-
DE 933R.
GN OrderedLocusNames=z2347;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
```

OC Enterobacteriaceae; Escherichia.
 RN [1]_TaxID=83334;
 RP SEQUENCE FROM N.A.
 RC STRAIN=015:H7 / EDL933 / ATCC 700927 / EHEC;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Ferris N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamoukis K.,
 RA Apodaca J., Anantharajan T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AF005367; AAG56394.1; -.
 DR PIR; F85741; F85741.
 DR PIR; H90768; H90768.
 DR HSP; P53636; 1E0W.
 DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00332; SOD_CU_ZN; 1.
 KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
 SQ SEQUENCE 274 AA; 28777 MW; 51DF1AFBF50989 CRC64;
 Query Match 43.6%; Score 414; DB 2; Length 274;
 Best Local Similarity 47.8%; Pred. No. 4.6e-29;
 Matches 88; Conservative 27; Mismatches 53; Indels 16; Gaps 5;
 QY 1 MKIKLFFVTSIVTSLTSITSVLACSVTSE-----VHMDDNGIKOSIGTIVTFTDCKG 56
 Db 100 MKKI-----AAATAMLT-----ASCCYAAEQEVPNMLVSADGKSVISGKITIQETPYG 149
 QY 57 LQIKTKLPGAGEHGFHIEGGSCGPAEHDGHLTAGLQAHGHYDPDKTGKHEGSLG-NG 115
 Db 150 LLFTPALHSLSEGIHGFVHEKNCALXGKPVAAISAGHFDPKNTGKHLPWSPDG 209
 QY 116 HKGDLPRLVVADGIKETLLAPRL-TVKEIKGTVMHAGGNYSDKPLPLGGGGARIA 174
 Db 210 HPGDLPALFVTHDGKANYPVLAPLNSLKEIKGRSLMLHAGGNDHHDHPEPLGGGARMA 269
 QY 175 CGVI 178
 Db 270 CGII 273
 RESULT 18
 SODC BRUAB STANDARD; PRT; 154 AA.
 AC P15453;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 05-JUN-2004 (Rel. 44, Last annotation update)
 DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
 GN Name=sodC;
 OS Brucella abortus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OC NCBI_TaxID=235;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=90148961; PubMed=2105741;
 RA Beck B.L., Tabatabai L.B., Mayfield J.E.;
 RT "A protein isolated from Brucella abortus is a Cu-Zn superoxide

dismutase";
 RL Biochemistry 29:372-376(1990).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR PIR; A33893; A33893.
 DR HSP; P53636; 1E0W.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN; 1.
 DR PROSITE; PS00332; SOD_CU_ZN; 1.
 KW Antioxidant; Copper; Direct protein sequencing; Metal-binding;
 KW Oxidoreductase; Periplasmic; Zinc.
 FT METAL 48 48 Copper (By similarity).
 FT METAL 50 50 Copper (By similarity).
 FT METAL 73 73 Copper and zinc (By similarity).
 FT METAL 82 82 Zinc (By similarity).
 FT METAL 90 90 Zinc (By similarity).
 FT METAL 93 93 Zinc (By similarity).
 FT METAL 128 128 Copper (By similarity).
 FT DISULFID 55 150 By similarity.
 SQ SEQUENCE 154 AA; 16072 MW; 4672C31481704468 CRC64;
 Query Match 43.5%; Score 413.5; DB 1; Length 154;
 Best Local Similarity 52.0%; Pred. No. 2.6e-29;
 Matches 79; Conservative 19; Mismatches 53; Indels 1; Gaps 1;
 QY 28 SVTSEVHMDDNGIKOSIGTIVTFTDCKGLQIKTKLPGAGEHGFHIEGGSCGPAEHD 87
 Db 2 SITVQMEALPTGPKKEVGTVISEAPGLHFKVNMEXLTPGYGHFVHNPSCAPEKD 61
 QY 88 GHLTAGLQAHGHYDPDKTGKHEGSLGNGHKGDLPRLVVADGIKETLLAPRL-TVKEIK 146
 Db 62 GKIVPALAAGGHYDPGNTHHLGPEGDGHMGDLPRLSANADGKVSFTVVAHLKLAEIK 121
 QY 147 GRTVMHAGGNYSDKPLPLGGGGARIA 178
 Db 122 QRLSMVHVGNDYSDKPEPLGGGARFACGVI 153
 RESULT 19
 Q9MCO2 PRELIMINARY; PRT; 174 AA.
 AC Q9MCO2;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Putative Cu/Zn superoxide dismutase.
 GN Name=sodCIII;
 OS Phage Fels-1.
 OC Viruses.
 OC NCBI_TaxID=128975;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21065122; PubMed=11136448;
 RA Figueroa-Bossi N., Uzzau S., Maloriol D., Bossi L.;
 RT "Variable assortment of prophages provides a transferable repertoire
 RT of pathogenic determinants in Salmonella".
 RL Mol. Microbiol. 39:260-271(2001).
 DR EMBL; AF254784; AAF62484.1; -.
 DR HSP; P53636; 1E0W.
 DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.

```
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
SQ SEQUENCE 174 AA; 17995 MW; 557DE7F2B5DB76D0 CRC64;

Query Match 43.5%; Score 413.5; DB 2; Length 174;
Best Local Similarity 49.4%; Pred. No. 3e-29;
Matches 84; Conservative 21; Mismatches 62; Indels 3; Gaps 3;

QY 11 IVTISLLTSITSVLACSVTSEVHMDDNGIKQSIGTFTDTDKGLQIKTDKGLPAGE 70
Db 5 LMTLAVFL-FSSAASAASTATVNLVANGTGQKIGNITITETEGYGLLFTPLHSLPAGI 63

QY 71 HGPHIHGGSCGPAEDHGHLTAGLQAHGHYDPDKTGKHEGPLG-NGHKDGLPRLVVKADG 129
Db 64 HGPHIHGGSCDAGMKDGPVAAALAGGHLDPQHTNKGHPYNPEGHGLGDLPALYVQAG 123

QY 130 IAKETILLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIACGVI 178
Db 124 NADYPVLAPRLKSIQVKGHAIMHTGDNHEDHPNPLGGGGARIACGVI 173

RESULT 20
Q8ZQF7 PRELIMINARY; PRT; 174 AA.
AC Q8ZQF7
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Putative Fels-1 prophage Cu/Zn superoxide dismutase (EC 1.15.1.1)
DE (Superoxide dismutase [Cu-Zn]).
GN OrderedLocusNames=STM0924;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
CC EMBL; AE008737; AAL19860.1; -.
DR HSSP; P53636; 1EQW.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; Sod_Cu; 1.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 174 AA; 17995 MW; 557DE7F2B5DB76D0 CRC64;

Query Match 43.5%; Score 413.5; DB 2; Length 174;
Best Local Similarity 49.4%; Pred. No. 3e-29;
Matches 84; Conservative 21; Mismatches 62; Indels 3; Gaps 3;

QY 11 IVTISLLTSITSVLACSVTSEVHMDDNGIKQSIGTFTDTDKGLQIKTDKGLPAGE 70
Db 5 LMTLAVFL-FSSAASAASTATVNLVANGTGQKIGNITITETEGYGLLFTPLHSLPAGI 63

QY 71 HGPHIHGGSCGPAEDHGHLTAGLQAHGHYDPDKTGKHEGPLG-NGHKDGLPRLVVKADG 129
Db 64 HGPHIHGGSCDAGMKDGPVAAALAGGHLDPQHTNKGHPYNPEGHGLGDLPALYVQAG 123

QY 130 IAKETILLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIACGVI 178
Db 124 NADYPVLAPRLKSIQVKGHAIMHTGDNHEDHPNPLGGGGARIACGVI 173

RESULT 20
Q8ZQF7 PRELIMINARY; PRT; 174 AA.
AC Q8ZQF7
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Putative Fels-1 prophage Cu/Zn superoxide dismutase (EC 1.15.1.1)
DE (Superoxide dismutase [Cu-Zn]).
GN OrderedLocusNames=STM0924;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
CC EMBL; AE008737; AAL19860.1; -.
DR HSSP; P53636; 1EQW.
DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006801; P:superoxide metabolism; IEA.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; Sod_Cu; 1.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 174 AA; 17995 MW; 557DE7F2B5DB76D0 CRC64;

Query Match 43.5%; Score 413.5; DB 2; Length 174;
Best Local Similarity 49.4%; Pred. No. 3e-29;
Matches 84; Conservative 21; Mismatches 62; Indels 3; Gaps 3;

QY 11 IVTISLLTSITSVLACSVTSEVHMDDNGIKQSIGTFTDTDKGLQIKTDKGLPAGE 70
Db 5 LMTLAVFL-FSSAASAASTATVNLVANGTGQKIGNITITETEGYGLLFTPLHSLPAGI 63

QY 71 HGPHIHGGSCGPAEDHGHLTAGLQAHGHYDPDKTGKHEGPLG-NGHKDGLPRLVVKADG 129
Db 64 HGPHIHGGSCDAGMKDGPVAAALAGGHLDPQHTNKGHPYNPEGHGLGDLPALYVQAG 123

QY 130 IAKETILLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIACGVI 178
Db 124 NADYPVLAPRLKSIQVKGHAIMHTGDNHEDHPNPLGGGGARIACGVI 173

RESULT 21
Q7AEAL PRELIMINARY; PRT; 175 AA.
AC Q7AEAL
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative copper/zinc-superoxide dismutase.
GN OrderedLocusNames=ECs1989;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL Nat. Res. 8:11-22(2001).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
CC EMBL; AF002556; BAB35412.1; -.
DR InterPro; IPR001424; SOD_CU_ZN.
DR Pfam; PF00080; Sod_Cu; 1.
DR ProDom; PD000469; SOD_CU_ZN; 1.
DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
KW Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 175 AA; 18275 MW; 1058B6A5F7B9478E CRC64;

Query Match 43.5%; Score 413; DB 2; Length 175;
Best Local Similarity 47.8%; Pred. No. 3.4e-29;
Matches 88; Conservative 27; Mismatches 53; Indels 16; Gaps 5;

QY 1 MKIKLFFVTSIVTISLLTSITSVLACSVTSE- - -VHMDDNGIKQSIGTFTDTDKG 56
Db 1 MKCKI- - - -IAAIAMLTA- - - -ASCGYAAEQEVPMLVLSADGKEVSIGKITIETPYG 50

QY 57 LOIKTDLKLPAHEHGFHIEGSCGPAEDHGHLTAGLQAHGHYDPDKTGKHEGPLG-NG 115
Db 51 LLFTPALHSLSEHGFHIEGSCGPAEDHGHLTAGLQAHGHYDPDKTGKHEGPLG-NG 110

QY 116 HGKGLPRLVVKADGIAKETILLAPRL-TVKEIKGRTVMIHAGGDNYSKPLPLGGGGARIA 174
Db 111 HLGDLPALFVTHDGKANYVPLAPRLNLSKEIKGRSLMLHAGGDNHHDHPEPLGGGGARIA 170

QY 175 CGVI 178
Db 171 CGII 174

RESULT 22
Q8X6B6 PRELIMINARY; PRT; 328 AA.
ID Q8X6B6
```

OC Q8X6B6;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Putative superoxide dismutase.
 GN OrderedLocusNames=z3312;
 OS Escherichia coli O157:H7;
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 ON NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDJ933 / ATCC 700927 / EHEC;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Probst E.J., Davis N.W., Lim A., Dimalanta E.F., Potamocousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AE005440; AAG57201.1; -;
 DR PIR; E85842; E85842.
 DR PIR; E90877; E90877.
 DR HSP; P53636.1E0W.
 DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR PROSITE; PS00332; SOD_CU_ZN; 2; 1.
 DR Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
 KW SEQUENCE 328 AA; 34217 MW; E3DE6E5C988E5A6 CRC64;
 QY 1 MKIKLFFVTSIVTSLTSTSVVLACSVTSE----VHMIDNDGKIKOSIGTVFTFTDCKG 56
 DB 154 MKCKI-----IAAIAMLTA-----ASCGYAAEOEVPMLNLSADGKEVSIKTIQETPYG 203
 QY 57 LQIKTDLKGIPAGBHGHIHEGGSCGPAEDHGLTAGLQAHGHYDPKTKGHEGPIG-NG 115
 DB 204 LLFTPALHLSSEGHGHFVHEKGNCPALPKDGFVAALSAGGHFDPNNTKGILGPSDPG 263
 QY 116 HKGDLPLRVKADGIKTELAPRL-TVKEIKGRTVMIHAGSDNYDKPLPLGGGGARIA 174
 DB 264 HLGDLPALFVTHDGKANYVPVLAPELNLKEIKGSLMLHAGGDNHHDHPEPLGGGGARWA 323
 QY 175 CGVI 178
 DB 324 CGII 327
 RESULT 23
 Q7MFW9
 ID Q7MFW9 PRELIMINARY; PRT; 170 AA.
 AC Q7MFW9;
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DE Superoxide dismutase, Cu-Zn.
 GN Name=VVA0291;
 OS Vibrio vulnificus (strain YJ016).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 ON NCBI_TaxID=196600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14658965;
 RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
 RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
 RA Shao C.-P., Lee C.-I., Hor L.-I., Tsai S.-F.;
 RT "Comparative genome analysis of Vibrio vulnificus, a marine
 RT pathogen";
 RL Genome Res. 13:2577-2587(2003).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 DR EMBL; AF005345; BAC96317.1; -;
 DR GO; GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0006801; P:superoxide metabolism; IEA.
 DR InterPro; IPR001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR PROSITE; PS000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN; 1; UNKNOWN 1.
 DR Copper; Metal-binding; Oxidoreductase; Zinc.
 KW SEQUENCE 170 AA; 17613 MW; 5AA2C93F1176704A CRC64;
 QY 10 SIVTSLTSTSVVLACSVTSEVHMIDNDGKIKOSIGTVFTDCKGLQIKTLKGLPAG 69
 DB 5 TLAAAILLYSTSSFAQSLSVD-----MKDLSSNQLTGTVTSSSDYGTGVTFFDLKGLPSG 59
 QY 70 EHGFIHEGGSCGPAEDHGLTAGLQAHGHYDPKTKGHEGP-LGNHGKGLPLRVKAD 128
 DB 60 LHGFHLHANGSCSSNKGDKVVLGGAGGHYDPNNSGKHGYPTWEDNHLGLPALFVDAS 119
 QY 129 GKAKETLLAPRLTVKEIKGRTVMIHAGSDNYDKPLPLGGGGARIACGVI 178
 DB 120 GNASQVLAPELNLKEIKGSLMLHAGGDNHHDHPEPLGGGGARIVCGVI 169
 Query Match 43.3%; Score 411; DB 2; Length 170;
 Best Local Similarity 48.2%; Pred. No. 4.9e-29;
 Matches 82; Conservative 26; Mismatches 56; Indels 6; Gaps 2;
 QY 10 SIVTSLTSTSVVLACSVTSEVHMIDNDGKIKOSIGTVFTDCKGLQIKTLKGLPAG 69
 DB 5 TLAAAILLYSTSSFAQSLSVD-----MKDLSSNQLTGTVTSSSDYGTGVTFFDLKGLPSG 59
 QY 70 EHGFIHEGGSCGPAEDHGLTAGLQAHGHYDPKTKGHEGP-LGNHGKGLPLRVKAD 128
 DB 60 LHGFHLHANGSCSSNKGDKVVLGGAGGHYDPNNSGKHGYPTWEDNHLGLPALFVDAS 119
 QY 129 GKAKETLLAPRLTVKEIKGRTVMIHAGSDNYDKPLPLGGGGARIACGVI 178
 DB 120 GNASQVLAPELNLKEIKGSLMLHAGGDNHHDHPEPLGGGGARIVCGVI 169
 RESULT 24
 Q87G06
 ID Q87G06 PRELIMINARY; PRT; 171 AA.
 AC Q87G06;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Superoxide dismutase, Cu-Zn.
 GN OrderedLocusNames=VPA1514;
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 ON NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIND 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Nejima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae";
 RL Lancet 361:743-749(2003).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems (By similarity).
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By


```

RESULT 27
Q8ZBN3 PRELIMINARY; PRT; 201 AA.
ID Q8ZBN3; Q74XS8;
AC AAS60586 PRELIMINARY; PRT; 201 AA.
DT 01-MAR-2004 (TrEMBLrel. 20, Created)
DT 01-MAR-2004 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 28, Last annotation update)
DE Superoxide dismutase (EC 1.15.1.1)
GN Name=sodC; OrderedLocusNames=YPO311, YPO3375;
OS Versinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.-M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston F.C.F., Quail M.A., Rutherford K.M.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Mediaevalis;
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
RA Yang R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS017128; AAS60586.1;
SQ SEQUENCE 201 AA; 20675 MW; 75D25ABE48BF8DCD CRC64;
Query Match 42.5%; Score 404; DB 2; Length 201;
Best Local Similarity 54.2%; Pred. No. 2.6e-28;
Matches 83; Conservative 18; Mismatches 50; Indels 2; Gaps 2;

QY 28 SVTSEVHMDDNGIKQSIGVTFTDQKGLQIKTDKGLPAGEHGPHIHEGSCGPAEHD 87
DB 48 SMTVKINESLPQNGKALGTVTVETAYGLTFPHLTGLAPGHGPHLHKPSCAPGMKD 107
QY 88 GHLTAGLQAHGYDPDKTGKEGFLGN-GHKGDPLRLVVKADGIAKETLLAPRL-TVKEI 145
DB 108 GKAVPALAAGGHLDPNKTGVHLGPYNDKGHLGDLPLGLVNVADGTATVPVLAFLKSLSEV 167
QY 146 KGRVTMIHAGDNYSDKPLPLGGGGARACGVI 178
DB 168 KQHALMIHAGDNYSDHPMPPLGGGGARMACGVI 200

RESULT 29
Q7VWVW9 PRELIMINARY; PRT; 170 AA.
ID Q7VWVW9
AC Q7VWVW9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative superoxide dismutase (EC 1.15.1.1).
GN OrderedLocusNames=BP2067;
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.B., Holden M.F.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders S., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).

```



```
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL: BX640417; CAE42345.1; -.
DR GO: GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO: GO:0046872; F:metal ion binding; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006801; P:superoxide metabolism; IEA.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; SOD_Cu; 1.
DR ProDom: PD000469; SOD_CU_ZN; 1.
DR PROSITE: PS00087; SOD_CU_ZN_1; UNKNOWN 1.
DR PROSITE: PS00332; SOD_CU_ZN_2; UNKNOWN 1.
KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 170 AA; 17422 MW; BFBC899F119BSA5E CRC64;

Query Match 42.0%; Score 399; DB 2; Length 170;
Best Local Similarity 47.6%; Pred. No. 6e-28;
Matches 81; Conservative 25; Mismatches 60; Indels 4; Gaps 3;

QY 10 SIVTISLLTSITSVLACSVTSEVHMIDDNGIKQSIGTFTTDTDKGLQIKTDLKGLPAG 69
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 3 SLIALA-LAAMSGVAAQAQSV--EMSLSPKGDQKIGQIKYEQTDYGVLFPTDLHGLAPG 59
QY 70 EHGPHIHGSGCPAEHDGHLTAGLQAHGHYDPDKTGKHEGPLG-NGHKGDLPRLVVKAD 128
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 60 VHGFLHDKPSCAMTKVDGKDMVGGAGGHWDPDKTGAHKGPYDKTGKGDLPVAVVAD 119
QY 129 GIAKETLLAPRLTVKEIKGRTVMIHAGDNYSDKPLPLGGGGARIACGVI 178
DQ ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 120 GKATYPLVAPRLKVLADLGHALMIHVGDNYSDHPAALGGGGGRMACGVV 169

RESULT 30
ID Q7WH53 PRELIMINARY; PRT; 170 AA.
AC Q7WH53;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative superoxide dismutase (EC 1.15.1.1).
GN OrderedLocusNames=BB3357;
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Harris D.E., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
RA Parkhill J., Sebahia M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagsels K.,
RA Leather S., Mouton S., Norbertzak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
DR EMBL: BX640447; CAE33849.1; -.
DR GO: GO:0004785; F:copper, zinc superoxide dismutase activity; IEA.
DR GO: GO:0046872; F:metal ion binding; IEA.
```

RX MEDLINE=97075068; PubMed=8917495;
 RA Bourne Y., Redford S.M., Steinman H.M., Lepock J.R., Tainer J.A.,
 RA Getzoff E.D.;
 RT "Novel dimeric interface and electrostatic recognition in bacterial
 RT Cu,Zn superoxide dismutase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:12774-12779(1996).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RX MEDLINE=99096923; PubMed=9878406;
 RA Bordo D., Matak D., Djinoovic-Carugo K., Rosano C., Pesce A.,
 RA Bolognesi M., Stroppolo M.E., Falconi M., Battistoni A., Desideri A.;
 RT "Evolutionary constraints for dimer formation in prokaryotic Cu,Zn
 RT superoxide dismutase.";
 RL J. Mol. Biol. 285:283-296(1999).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COPACITOR: Binds 1 copper ion and 1 zinc ion per subunit.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; J02658; AAA25632.1; -;
 DR PIR; A26689; DSFOCL.
 DR PDB; 1BZO; X-ray; A=23-173.
 DR PDB; 1IB5; X-ray; A=23-173.
 DR PDB; 1IB6; X-ray; A=23-173.
 DR PDB; 1IB7; X-ray; A=23-173.
 DR PDB; 1IBD; X-ray; A=23-173.
 DR PDB; 1IBF; X-ray; A=23-173.
 DR PDB; 1IBH; X-ray; A=23-173.
 DR PDB; 1OAJ; X-ray; A=23-173.
 DR PDB; 1OAL; X-ray; A=23-173.
 DR PDB; 1VAI; X-ray; A/B/C=23-173.
 DR InterPro: IP001424; SOD_CU_ZN.
 DR Pfam; PF00080; SOD_Cu; 1.
 DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN_1; 1.
 DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
 KW 3D-structure; Antioxidant; Copper; Direct protein sequencing;
 KW Metal-binding; Oxidoreductase; Periplasmic; Signal; Zinc.
 FT SIGNAL 1 23
 FT CHAIN 23 173 Superoxide dismutase [Cu-Zn].
 FT METAL 67 67 Copper.
 FT METAL 69 69 Copper.
 FT METAL 92 92 Copper and zinc.
 FT METAL 101 101 Zinc.
 FT METAL 110 110 Zinc.
 FT METAL 113 113 Zinc.
 FT METAL 147 147 Copper.
 FT DISULFID 74 169
 FT STRAND 24 31
 FT TURN 32 34
 FT STRAND 37 46
 FT TURN 47 48
 FT STRAND 49 56
 FT STRAND 62 64
 FT STRAND 66 69
 FT STRAND 77 79
 FT TURN 80 81
 FT STRAND 82 84
 FT TURN 85 86
 FT HELIX 87 89
 FT STRAND 92 92
 FT STRAND 95 96
 FT TURN 104 105

FT TURN 111 112
 FT STRAND 113 113
 FT STRAND 117 119
 FT TURN 121 122
 FT STRAND 125 125
 FT STRAND 129 131
 FT TURN 132 133
 FT HELIX 136 139
 FT TURN 140 141
 FT STRAND 143 147
 FT HELIX 159 162
 FT STRAND 166 171
 SQ SEQUENCE 173 AA; 18109 MW; 5931576F1E2ABF47 CRC64;
 Query Match 41.2%; Score 391; DB 1; Length 173;
 Best Local Similarity 47.4%; Pred. No. 3.2e-27;
 Matches 81; Conservative 26; Mismatches 58; Indels 6; Gaps 4;
 QY 10 SIVTISLTSTSVVLACSVTSEVMID-DNGIKOSIGTIVTFTDDKGLQIKTDIKGLPA 68
 Db 6 TLLFTALAFGLSHQALQDLT--VMTDLQTS--KPVGTIELSQNKYGVFTPELADLTP 61
 QY 69 GSHGFHIEHGSCGPAEDHGLTAGLQAHGHDYDPDKTGHGEP-LGNHGKGLDLRLVVKA 127
 Db 62 GWHGFHIEHGSCGPAEDHGLTAGLQAHGHDYDPDKTGHGEP-LGNHGKGLDLRLVVKA 121
 QY 128 DGIAKETLAPRLTVKTKGRTVMHAGDNYSDKPLPLGGGARIACGVI 178
 Db 122 NGLATNPVLAFLRLTKELKGHAIMIHAGDNDHSDMPRALGGGGARVACGVI 172
 RESULT 32
 SODC_HAEP
 ID SODC_HAEP STANDARD; PRT; 187 AA.
 AC P25842;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
 GN Name=sodC;
 OS Haemophilus parainfluenzae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 CC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=729;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1391;
 RX MEDLINE=92041655; PubMed=1938942;
 RA Kroll J.S., Langford P.R., Loynds B.M.;
 RT "Copper-zinc superoxide dismutase of Haemophilus influenzae and H.
 RT J. Bacteriol. 173:7449-7457(1991).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -!- FUNCTION: May confer survival advantage by accelerating
 CC dismutation of superoxide of environmental origin to hydrogen
 CC peroxide, disruptive to the normal mucociliary clearance process
 CC in the host.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COPACITOR: Binds 1 copper ion and 1 zinc ion per subunit (By
 CC similarity).
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

[illegible]

KW PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN 1.
KW PROSITE; PS00332; SOD_CU_ZN_2; UNKNOWN 1.
KW Complete proteome; Copper; Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 170 AA; 17871 MW; 5CB62830C4C2DE04 CRC64;

Query Match
Best Local Similarity 40.7%; Score 387; DB 2; Length 170;
Matches 85; Conservative 18; Mismatches 54; Indels 16; Gaps 4;

QY 14 ISLLTSIT-----SVWLACSVTSEVHMIDNGIKQSIGTVFTDTKGLQIKTLKGLP 67
DQ :
DB 5 ISLLIWLFLSFPFSLANAEIVAVNMVRVAKTHGSSVGVLTLDYTGILLIPNLRLGP 64

QY 68 AGE-HGFPHIEGGSCGPAEHGHILTAGLQAQHGYDPDKTKHGEPLG-NGHKGDLPRLVV 125
DQ :
DB 65 PGRIGHGFHIHVNPDCS-----NKGEAAGCHLDLPANTKGHLGPNPNHGDLGDPALTV 116

QY 126 KADGIATKTLLAPRLTVKEIKGRVTWMIHAGDNYSKDPLPLGGGGARIACGVI 178
DQ :
DB 117 EKNGGTATLPIAPLRKVNLGHSIMIHAGDNYSHPKELGGGGARIACGIV 169

RESULT 35
SODC_NEIMA
ID SODC_NEIMA STANDARD; PRT; 186 AA.
AC PS7005;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN Name=sodC; OrderedLocNames=NMA1617;
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RP MEDLINE=2022556; PubMed=10761919; DOI=10.1038/35006655;
RX Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.M.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.K., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagels K., Leather S., Moule S., Murgall K.L., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491".
RL Nature 404:502-506(2000).

-!- FUNCTION: Destroys radicals which are normally produced within the
cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
similarity).
CC
CC -!- SUBUNIT: Homodimer (By similarity).
CC
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC
CC -!- SIMILARITY: Belongs to the Cu-zn superoxide dismutase family.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

EMBL; AL162756; CAB84845.1; --
PIR; B81855; E81855.
HSP; P24702; 2APS.
InterPro; IPR001424; SOD_CU_ZN.
Pfam; PF00080; Sod_Cu_1--
ProDom; PD000469; SOD_CU_ZN; 1.
PROSITE; PS00087; SOD_CU_ZN_1; 1.
PROSITE; PS00332; SOD_CU_ZN_2; 1.
PROSITE; PS00087; SOD_CU_ZN_1; 1.

KW Antioxidant; Complete proteome; Copper; Metal-binding; Oxidoreductase;
KW Periplasmic; Signal; Zinc.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 186 Superoxide dismutase [Cu-Zn].
FT METAL 79 79 Copper (By similarity).
FT METAL 81 81 Copper (By similarity).
FT METAL 104 104 Copper and zinc (By similarity).
FT METAL 113 113 Zinc (By similarity).
FT METAL 122 122 Zinc (By similarity).
FT METAL 125 125 Zinc (By similarity).
FT METAL 160 160 Copper (By similarity).
FT DISULFID 86 182 By similarity.
SQ SEQUENCE 186 AA; 19550 MW; 04C84AB87FC7522D CRC64;

Query Match
Best Local Similarity 40.7%; Score 387; DB 1; Length 186;
Matches 79; Conservative 31; Mismatches 61; Indels 4; Gaps 3;

QY 8 VTSIVTISLSTS---ITSVLACSVTSEVHMIDNGIKQSIGTVFTDTKGLQIKTLKGLP 65
DQ :
DB 11 VSVCVSVAQAHEHNTIPKGSIEVKVQLDPVNGKNVDGTVTTITESNYGLVFPTDLQG 70

QY 66 LPAGEHGFHIHEGGSCGPAEHGHILTAGLQAQHGYDPDKTKHGEPLG-NGHKGDLPRLV 124
DQ :
DB 71 LSEGLAGFHIEHNPSCEPKKEKLTAGLCAGHWDPKAKQKHGWDDAHLGDPALT 130

QY 125 VKADGIATKTLLAPRLTVKEIKGRVTWMIHAGDNYSKDPLPLGGGGARIACGVI 178
DQ :
DB 131 VLHDGTATNEVLAPRLKLDDVRGHSIMIHAGDNYSHPKELGGGGARIACGVI 165

RESULT 36
SODC_NEIMB
ID SODC_NEIMB STANDARD; PRT; 186 AA.
AC Q59623;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN Name=sodC; OrderedLocNames=NME1398;
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=MC58 / Serogroup B;
RX MEDLINE=98084476; PubMed=9423860;
RA Walks K.E., Dunn K.L., Farrant J.L., Reddin K.M., Gorringe A.R.,
RA Langford P.R., Kröll J.S.;
RT "Periplasmic superoxide dismutase in meningococcal pathogenicity";
RL Infect. Immun. 66:213-217(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C.,
RA Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F.,
RA Dodson R.J., Nelson W.C., Quinn M.L., DeBoy R.T., Peterson J.D.,
RA Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
RA Dougherty B.A., Mason T.M., Ciecko A., Parksey D.S., Blair E.,
RA Cittono H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
RA Qin H., Vanathavan J.J., Gill J., Scarlato V., Masignani V., Pizza M.,
RA Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
RA Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58".
RL Science 287:1809-1815(2000).
RN [3]
RC SEQUENCE OF 87-173 FROM N.A.
RP STRAIN=MC58 / Serogroup B;
RX MEDLINE=96118708; PubMed=7496539;
RA Kröll J.S., Langford P.R., Wilks K.E., Keil A.D.

RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
RL from the eukaryotic enzyme, and not so rare after all!";
RC
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ001313; CAA04674.1; -.
CC EMBL; AB002488; AAF41762.1; -.
CC EMBL; X83126; CAA58207.1; -.
CC F1R; F81088; F81088.
CC HSSP; P24702; 2APS.
CC TIGR; NMB1398; -.
CC InterPro; IPR001424; SOD_CU_ZN.
CC Pfam; PF00080; SOD_CU_ZN; 1.
CC ProDom; PD000469; SOD_CU_ZN; 1.
CC PROSITE; PS00087; SOD_CU_ZN; 1.
CC PROSITE; PS00332; SOD_CU_ZN; 1.
CC Antioxidant; Complete proteome; Copper; Metal-binding; Oxidoreductase;
KW Periplasmic; Signal; Zinc.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 186 Superoxide dismutase [Cu-Zn].
FT METAL 79 79 Copper (By similarity).
FT METAL 81 81 Copper (By similarity).
FT METAL 104 104 Copper and zinc (By similarity).
FT METAL 113 113 Zinc (By similarity).
FT METAL 122 122 Zinc (By similarity).
FT METAL 125 125 Zinc (By similarity).
FT METAL 160 160 Copper (By similarity).
FT METAL 186 182 By similarity.
FT DISULFID 86 182
FT CONFLICT 130 130
FT CONFLICT 130 130
SQ SEQUENCE 186 AA; 19520 MW; 6499049BFA3427C CRC64;
Query Match 40.5%; Score 385; DB 1; Length 186;
Best Local Similarity 50.3%; Pred. No. 1.2e-26;
Matches 77; Conservative 22; Mismatches 52; Indels 2; Gaps 2;
Qy 28 SVTSEVMIDNGIKQSIGVTFTDTKGLQIKTDLKGHPAGHGHGSGCGPAEHD 87
Db 33 SIEYKQQLDPVNGKDVGVTTITSNYGLVFTPDQLGSEGLHGHFIHENPCEPKE 92
Qy 88 GHLTAGLQAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPLTL-VKEI 145
Db 93 GKLTAGLQAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPLTL-VKEI 152
Qy 146 KGRVTMIHAGDNYSDKPLPLGGGGARIACGVI 178
Db 153 RGHSTIMHTGDNHSDHPAPLPGGGGPRMACGVI 185
RESULT 37
SODC_PASMU STANDARD; PRT; 186 AA.
AC Q59689;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).
GN Name=sodC; Ordered locus names=PM1952;
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

Pasteurellaceae; Pasteurella.
NCBI_TaxID=747;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=PM70.
RA MEDLINE=11145866; PubMed=11248100; DOI=10.1073/pnas.051634598;
RA May B.J., Zhang Q., Li L.B., Paustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida PM70.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
RL [2]
RN SEQUENCE OF 87-173 FROM N.A.
RP STRAIN=TS.
RC MEDLINE=96118708; PubMed=7496539;
RA Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
RA "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
RT from the eukaryotic enzyme, and not so rare after all!";
RL Microbiology 141:2271-2279(1995).
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems (By similarity).
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB006232; AAK04036.1; -.
CC EMBL; X83124; CAA58205.1; -.
CC HSSP; P24702; 2APS.
CC InterPro; IPR001424; SOD_CU_ZN.
CC Pfam; PF00080; SOD_CU_ZN; 1.
CC ProDom; PD000469; SOD_CU_ZN; 1.
CC PROSITE; PS00087; SOD_CU_ZN; 1.
CC PROSITE; PS00332; SOD_CU_ZN; 1.
CC Antioxidant; Complete proteome; Copper; Metal-binding; Oxidoreductase;
KW Periplasmic; Signal; Zinc.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 186 Superoxide dismutase [Cu-Zn].
FT METAL 79 79 Copper (By similarity).
FT METAL 81 81 Copper (By similarity).
FT METAL 104 104 Copper and zinc (By similarity).
FT METAL 113 113 Zinc (By similarity).
FT METAL 122 122 Zinc (By similarity).
FT METAL 125 125 Zinc (By similarity).
FT METAL 160 160 Copper (By similarity).
FT DISULFID 86 182
FT CONFLICT 130 130
FT CONFLICT 130 130
SQ SEQUENCE 186 AA; 19718 MW; 34D702070EE0DDC CRC64;
Query Match 40.2%; Score 382; DB 1; Length 186;
Best Local Similarity 51.7%; Pred. No. 2.3e-26;
Matches 77; Conservative 17; Mismatches 53; Indels 2; Gaps 2;
Qy 32 EVHMDNDNGIKQSIGVTFTDTKGLQIKTDLKGHPAGHGHGSGCGPAEHDGHLT 91
Db 37 KVELLDPAKGNQIGKGVITSEYGLVFTPELKNITAGLHGHFIHQNPSCDAKEDGKLV 96
Qy 92 AGLQAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPLTL-VKEIKGRT 149
Db 97 AGLQAGHYDPDKTGKHEGP-LGNHGKGLDPLRLVWKADGIAKETLLAPLTL-VKEIKGRT 156
Qy 150 VMIHAGDNYSDKPLPLGGGGARIACGVI 178
Db 157 LMIHEGDNHSDHPAPLPGGGGPRMACGVI 185

RESULT 38
 SODC ACTPL STANDARD; PRT; 190 AA.
 AC P24702; 059135;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Superoxide dismutase [Cu-Zn] precursor (BC 1.15.1.1).
 GN Name=sodC;
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Actinobacillus.
 OX NCBI_TaxID=715;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Serotype III / Isolate 1421 (Nielsen);
 RX MEDLINE=97101016; PubMed=8945543;
 RA Langford P.R., Loyds B.M., Kroll J.S.;
 RT "Cloning and molecular characterization of Cu,Zn superoxide dismutase
 from Actinobacillus pleuropneumoniae.";
 RL Infect. Immun. 64:5035-5041(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S 4074 / Serotype 1;
 RA Helie M.C., Sirois M., Ouellet C., Boissinot M.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 180-190 FROM N.A.
 RC STRAIN=Serotype III / Isolate 1421 (Nielsen);
 RX MEDLINE=92158680; PubMed=1741300;
 RA Loyds B.M., Langford P.R., Kroll J.S.;
 RT "recF in Actinobacillus pleuropneumoniae.";
 RL Nucleic Acids Res. 20:615-615(1992).
 RN [4]
 RP SEQUENCE OF 91-177 FROM N.A.
 RC STRAIN=Serotype III / Isolate 1421 (Nielsen);
 RX MEDLINE=96118708; PubMed=7496539;
 RA Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;
 RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct
 from the eukaryotic enzyme, and not so rare after all!";
 RL Microbiology 141:2271-2279(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=20124004; PubMed=10656823;
 RA Forest K.T., Langford P.R., Kroll J.S., Getzoff E.D.;
 RT "Cu,Zn superoxide dismutase structure from a microbial pathogen
 establishes a class with a conserved dimer interface.";
 RL J. Mol. Biol. 296:145-153(2000).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X99396; CAA6771.1; -;
 CC EMBL; U51440; AAB02816.1; -;
 CC EMBL; X63626; CAA45174.1; -;
 CC EMBL; X82123; CAA58204.1; -;
 CC F01; I39650; I39650.
 CC DB; 2APS; X-ray; A/B=29-190.
 CC InterPro; IPR001424; SOD_CU_ZN.
 CC Pfam; PF00080; Sod_Cu_1.
 CC

DR ProDom; PD000469; SOD_CU_ZN; 1.
 DR PROSITE; PS00087; SOD_CU_ZN_1; 1.
 DR PROSITE; PS00332; SOD_CU_ZN_2; 1.
 KW 3D-structure; Antioxidant; Copper; Metal-binding; Oxidoreductase;
 KW Periplasmic; Signal; Zinc.
 FT SIGNAL 1 23
 FT CHAIN 24 190 Superoxide dismutase [Cu-Zn].
 FT METAL 83 83 Copper.
 FT METAL 85 85 Copper.
 FT METAL 108 108 Copper and zinc.
 FT METAL 117 117 Zinc.
 FT METAL 126 126 Zinc.
 FT METAL 129 129 Zinc.
 FT METAL 164 164 Copper.
 FT DISULFID 90 186
 FT CONFLICT 113 113 E -> D (in Ref. 2).
 FT CONFLICT 124 124 N -> D (in Ref. 2).
 FT CONFLICT 134 136 FVE -> TIA (in Ref. 2).
 FT STRAND 38 44
 FT TURN 47 49
 FT STRAND 52 62
 FT TURN 63 64
 FT STRAND 65 72
 FT STRAND 78 80
 FT STRAND 82 85
 FT STRAND 93 95
 FT TURN 96 97
 FT STRAND 98 100
 FT TURN 101 102
 FT HELIX 103 105
 FT STRAND 108 108
 FT TURN 111 112
 FT TURN 120 121
 FT TURN 123 124
 FT TURN 127 128
 FT STRAND 129 129
 FT TURN 133 135
 FT TURN 137 138
 FT STRAND 141 141
 FT STRAND 145 147
 FT TURN 148 149
 FT HELIX 153 156
 FT TURN 157 158
 FT STRAND 159 164
 FT HELIX 176 179
 FT STRAND 183 190
 SQ SEQUENCE 190 AA; 20207 MW; 5FC1F3148972DC83 CRC64;
 Query Match 39.6%; Score 376; DB 1; Length 190;
 Best Local Similarity 51.7%; Pred. No. 9e-26;
 Matches 77; Conservative 18; Mismatches 52; Indels 2; Gaps 2;
 QY 32 EVHMDNGIKQSIGTFTTDTKGLQIKTDKGLPAGEHGFHIEGSGCPAEHDGHLT 91
 DB 41 QVQQLDPVKGNKDVGTVEITESAVGLVTFPHLHGLAQGLHGFHQNPSCEPKDKGLV 100
 QY 92 AGLQAHGHYDPDKTKEHGLPLG-NGHKGDLPLRVVVKADGIKATLLAPRL-TVKIKGRT 149
 DB 101 AGLGAGGHVDFPKTKQHGYPMSDNALHGLDLPALFVEHDSATNPLAPRLKLDVKVGH 160
 QY 150 VMIHAGDNTSDKPLPLGGGGARIACGVI 178
 DB 161 LMIHEGDNHSDHAPLPLGGGPRMACGVI 189
 RESULT 39
 Q6G3C6
 ID Q6G3C6 PRELIMINARY; PRT; 175 AA.
 AC Q6G3C6;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Superoxide dismutase.

Db 111 GLGAGCHNDPKQTKHGYEWSDDAHMGDLPALFVNHGGSATTPVLAAPRLKKLAEVKGHSL 170
QY 151 MIHAGGDNYSDKPLPLGGGGARIACGVI 178
Db 171 MIHAGGDNHSDEHPAPLGGGGPRMACGVI 198

Search completed: October 26, 2004, 09:44:55
Job time : 198 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2004, 00:05:50 ; Search time 92 seconds
(without alignments)
3878.435 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMGSLFICATGKTHSTGLG.....NSKSVTTADTMOQLALELKR 502

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 353394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10009823/runat_26102004_100130_1634/app_query.fasta_1.647
-DB=Issued Patents NA -QFWT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10009823@cgn 1.1.54 @runat_26102004_100130_1634 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEX=7

Database : Issued Patents NA:

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	528.5	20.4	1416	US-09-252-991A-2955	Sequence 2955, Ap
2	528.5	20.4	1512	US-09-252-991A-2728	Sequence 2728, Ap
3	475.5	18.4	1239	US-09-543-681A-4010	Sequence 4010, Ap
4	414.5	16.0	64681	US-09-790-988-1	Sequence 1, Appli
5	307	11.9	810	US-09-252-991A-3144	Sequence 3144, Ap
6	302	11.7	1545	US-09-252-991A-2823	Sequence 2823, Ap
7	286	11.1	798	US-09-543-681A-3938	Sequence 3938, Ap
8	281.5	10.9	1800	US-08-483-857-1	Sequence 1, Appli
9	277.5	10.7	1800	US-08-436-748-1	Sequence 1, Appli
10	237.5	9.2	789	US-08-483-857-3	Sequence 3, Appli
11	223.5	8.6	789	US-08-436-748-4	Sequence 4, Appli
12	193	7.5	5445	US-09-689-065B-2	Sequence 2, Appli

c	13	176	6.8	3129	4	US-09-252-991A-13873	Sequence 13873, A
	14	176	6.8	8211	4	US-09-252-991A-13656	Sequence 13656, A
	15	174	6.7	408	4	US-09-252-991A-3047	Sequence 3047, Ap
	16	173.5	6.7	3285	4	US-09-206-942-68	Sequence 68, Appl
	17	173.5	6.7	5116	1	US-08-302-832-1	Sequence 1, Appli
	18	173.5	6.7	5116	2	US-08-302-832-1	Sequence 1, Appli
	19	173.5	6.7	5116	2	US-08-530-198-1	Sequence 1, Appli
	20	173.5	6.7	5116	2	US-08-459-880-1	Sequence 1, Appli
	21	173.5	6.7	5116	2	US-08-728-470-1	Sequence 1, Appli
	22	173.5	6.7	5116	2	US-08-617-697-1	Sequence 1, Appli
	23	173.5	6.7	5116	3	US-08-719-641-1	Sequence 1, Appli
	24	173.5	6.7	5116	4	US-09-206-942-66	Sequence 66, Appl
	25	173.5	6.7	9171	1	US-08-038-682-5	Sequence 5, Appli
	26	173.5	6.7	9171	1	US-08-302-832-5	Sequence 5, Appli
	27	173.5	6.7	9171	2	US-08-530-198-5	Sequence 5, Appli
	28	173.5	6.7	9171	2	US-08-469-880-5	Sequence 5, Appli
	29	173.5	6.7	9171	2	US-08-728-470-5	Sequence 5, Appli
	30	173.5	6.7	9171	2	US-08-617-697-5	Sequence 5, Appli
	31	173.5	6.7	9171	3	US-08-719-641-5	Sequence 5, Appli
	32	165	6.4	25165	3	US-09-453-702B-39	Sequence 39, Appl
	33	160	6.2	4188	4	US-09-252-991A-13774	Sequence 13774, A
	34	155	6.0	810	1	US-08-436-748-2	Sequence 2, Appli
	35	155	6.0	810	3	US-08-483-857-2	Sequence 2, Appli
	36	151.5	5.9	4403765	3	US-09-103-840A-2	Sequence 1, Appli
	37	151.5	5.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	38	149.5	5.8	3543	4	US-09-206-942-64	Sequence 64, Appl
	39	149.5	5.8	3568	4	US-09-206-942-62	Sequence 62, Appl
	40	145.5	5.6	1284	4	US-09-252-991A-2729	Sequence 2729, Ap
	41	144	5.6	2831	4	US-09-477-135A-117	Sequence 117, App
	42	144	5.6	2852	3	US-09-056-556-203	Sequence 203, App
	43	144	5.6	2852	4	US-09-072-596-198	Sequence 198, App
	44	144	5.6	2852	4	US-09-072-967-203	Sequence 203, App
	45	142	5.5	3288	4	US-09-206-942-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1

US-09-252-991A-2955
; Sequence 2955, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2955

; LENGTH: 1416

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-2955

Alignment Scores:			
Pred. No.:	4,66e-47	Length:	1416
Score:	528.50	Matches:	151
Percent Similarity:	45.84%	Conservative:	86
Best Local Similarity:	29.21%	Mismatches:	205
Query Match:	20.44%	Indels:	75
DB:	4	Gaps:	16

US-10-009-823A-1 (1-502) x US-09-252-991A-2955 (1-1416)

Qy 4 SerLeuPheLeGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23

Db 31 AGTTTCACATCGCGCTGAGCGGCCTAGCGGCTGACGTACCGCGC 90

Qy	24	Asn	Asn	Ile	Ala	Asn	Ala	Asn	Thr	Ile	Gly	Tyr	Lys	Gln	Gln	Ile	Val	Val	Phe	Gln	Asp	43
Db	91	AA	CA	CA	TC	GC	CA	CG	CG	CA	CC	GC	CA	CG	CG	CA	TC	CA	AG	CA	AT	150
Qy	44	Leu	Phe	Ser	Gln	Asp	Leu	Ala	Ile	Gly	Ser	Thr	Gly	Ser	Gln	Gly	Pro	Asn	Gln	Ala	Gly	63
Db	151	GT	CT	AG	CG	CG	CT	CG	GT	---	CT	GG	TT	CG	---	GG	CAG	CA	AC	CC	CG	195
Qy	64	Met	Gly	Ala	Gln	Val	Gly	Ser	Val	Arg	Thr	Ile	Phe	Thr	Gln	Gly	Ala	Phe	Gly	Pro	Gly	83
Db	196	AG	CG	CG	GT	CT	CT	CG	CA	GT	CT	CG	CAG	AT	GT	TC	AA	AG	CA	GG	CA	255
Qy	84	Asn	Ser	Val	Thr	Asp	Leu	Ala	Ile	Gly	Gly	Lys	Gly	Phe	Phe	Gln	Val	Thr	Leu	Glu	Asp	103
Db	256	AA	CA	CG	GT	GT	GA	CA	CT	GG	CA	TC	AA	CG	CA	CG	CT	TC	TC	TC	CA	315
Qy	104	Lys	Val	His	Tyr	Thr	Arg	Ala	Gly	Asn	Phe	Arg	Phe	Thr	Gln	Asp	Gly	Phe	Leu	Asn	Asp	123
Db	316	CG	CA	TC	AG	CT	AC	CC	CG	CG	CT	AC	AT	CA	CC	GA	CA	AG	CA	GG	AT	375
Qy	124	Pro	Ser	Gly	Phe	Thr	Leu	Met	Gly	Ser	Arg	Ile	Ser	Asn	Asn	Pro	Asn	Ile	Lys	Glu	143	
Db	376	AA	CA	CG	CT	AC	CG	CT	CA	CG	GT	AT	GC	CG	CG	CA	CG	CA	CG	CA	CG	435
Qy	144	Thr	Leu	Glu	Pro	Ile	Gln	Leu	Asp	---	Phe	Asn	Asp	Pro	Thr	Val	Ala	Lys	Ser	Pro	Ala	162
Db	436	GT	GG	TC	AC	CG	CA	CT	CA	AG	TC	AG	CG	CG	CA	AT	CAG	CG	CG	CA	CG	495
Qy	163	Lys	Thr	Ser	Thr	Ala	Leu	Asn	Ala	Val	Val	Asn	Leu	Gly	Asp	Ser	Thr	Asp	Lys	Thr	Gln	182
Db	496	CA	CA	GT	CG	TC	AA	CA	CT	CA	AC	TC	CA	CG	CT	GA	AG	---	CCG	531		
Qy	183	Ser	Glu	Ala	Asn	Pro	Tyr	Phe	Ala	Leu	Glu	Ser	Trp	Lys	Gly	Asn	Gly	Thr	Pro	Pro	202	
Db	532	CC	GA	CC	GT	GA	CG	CG	CT	TC	---	---	---	---	---	---	---	---	---	---	555	
Qy	203	Ile	Ser	Thr	Ser	Asn	Tyr	Ser	Tyr	Ala	Gln	Pro	Met	Arg	Val	Tyr	Asp	Gln	Gly	Asn	222	
Db	556	TCC	AC	AC	CG	CG	TAC	CT	CA	AC	TC	CT	CT	CT	CG	GG	CA	TC	AT	GA	CG	615
Qy	223	Ser	His	Asp	Ile	Thr	Val	Tyr	Phe	---	---	---	---	---	---	---	---	---	---	---	237	
Db	616	TCC	CA	CA	CA	CA	AG	CA	AG	TC	CT	CT	CA	GA	CA	CG	CG	CA	CG	CA	CG	675
Qy	238	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	735	
Db	676	ATT	CCG	GAG	AAC	AG	CT	GA	AT	GCT	GA	AT	CG	AC	GG	CG	TC	AT	CC	GC	TC	771
Qy	254	Gly	Ser	Ala	Ala	Ser	Gly	Thr	Asp	Ser	Ala	Gly	Leu	Met	Ser	Gly	Thr	Met	Phe	273		
Db	736	TC	GA	CA	AA	GA	CG	CG	---	---	---	---	---	---	---	---	---	---	---	---	771	
Qy	274	Ser	Ser	Asn	Gly	Glu	Leu	Lys	Asn	Met	Thr	Ala	Phe	Thr	Pro	Thr	Gly	Ser	---	---	290	
Db	772	GA	CG	CA	CG	CG	CG	CA	GA	TC	CG	TT	CG	GG	CG	---	---	---	---	---	825	
Qy																						

```

Db      1015  TCCACCGCGTTCGCCAGAGCAACCCGATC-----CAGGACGGCTACACCACC 1062
Qy      385  GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
Db      1063  GGTGAGCTGGCAGCGCTGGAAATCGAGCACACGGGGTGATCTCCCGCTACACCAAT 1122
Qy      405  SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
Db      1123  GGCCAGTCCAAAGTGCAGGGCCAGCTGCTGGCGAACTTCGCCCAACATCCAGGSCCTG 1182
Qy      425  ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyClyProGluPheGly 444
Db      1183  ACGCCGATCGCGAAGACCTCCTGGGTGCAGTCCGCGGAGTCCGGCAGCGCGGTGCGC 1242
Qy      445  LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
Db      1243  CGCGCGCGCTCGCGGCACCTCGGGGGCGTTGCAATCCGGCGCGCTGGAGCGCTCCAACGTG 1302
Qy      465  AspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSer 484
Db      1303  GACATCTCCAACTGATGGTAACCTCATCTGTCACACGCGCAACTACCAGGCCAATGCC 1362
Qy      485  LysSerValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db      1363  AAGACCATCCAGACCGAGGATGCGGTGAGCCAGACCATCATCACTGCGC 1413

RESULT 2
US-09-252-991A-2728/c
; Sequence 2728, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2728
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2728

Alignment Scores:
Pred. No.: 5,19e-47 Length: 1512
Score: 528.50 Matches: 151
Percent Similarity: 45.84% Conservative: 86
Best Local Similarity: 29.21% Mismatches: 205
Query Match: 20.44% Indels: 75
DB: 4 Gaps: 16

US-10-009-823A-1 (1-502) x US-09-252-991A-2728 (1-1512)

```

5

Db 1338 ACGGCGTGTCTCTCTCGACGCTTCGAGATGTCAGAGGCAACATCGACTCGACC 1279
QY 84 AsnSerValThrAspLeuAlaIleGlyLysGlyPheGlnValThrLeuGluasp 103
Db 1278 AACACGCGTGTGACCTGACCAACGCAACGCGTCTTCCTCCACGCAACAAACGGG 1219
QY 104 LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAsp 123
Db 1218 GCGATCAGCTACACCGCGCGCGCTACTTCAATCCGACAGCAGGAGTTTCATCGTCGAC 1159
QY 124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGlu 143
Db 1158 AACACGCGTGTGACCTGACCAACGCAACGCGTCTTCCTCCACGCAACAAACGGG 1099
QY 144 ThrLeuGluProIleGlnLeuAsp---PheAsnAspProThrValAlaLysSerProAla 162
Db 1098 GTGGTCACCGACCTCAAGTCAGCGCGCCCAATCAGCGCGCGCAGGCACTCGAGCATC 1039
QY 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGln 182
Db 1038 CAGCAGTGTGTACACCTCACTCGACGCTGAAG-----CCG 1003
QY 183 SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPro 202
Db 1002 CCGACCGTGTGCGCGCTTC-----GATCGG 979
QY 203 IleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlyAsn 222
Db 978 TCCGACGCGGTACCTACCACTCGCTCTTCCTCGCGCATCTATGACAGCGGCAAC 919
QY 223 SerHisAspIleThrValTyrPhe-----AspGlyAlaProSerSerThr----- 237
Db 918 TCCACACATGAGCAGCAGTCTTTCATCAAGAACGCGCGGCGCAATGCGACCGCGCG 859
QY 238 -----GlySerLysThrPheGluTyrLeuVal---AlaMetAsnProSerGluAsp 253
Db 858 ATTCGCGAGAACAGCTGACCATGAAAGTGTGATCGACGCGCTCAATCGCTCGATCGG 799
QY 284 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuMetSerGlyThrMetThrPhe 273
Db 798 TCGAACAGACGCGCG-----ATGAGCTTCAACGTCACCTTC 763
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySer----- 290
Db 762 GACGCGACGCGCAGTACCTCGTTCGCGCG-----CGGACGCGCAGCAGCGCGG 709
QY 291 -----AlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuVal 304
Db 708 CCGGCGCTTCAGATCGACGCGCAGCAGCAGCAGTGTATCGATTCAGTCCGCGC----- 658
QY 305 AsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeu 324
Db 657 ACCGCGCAATCCGCGACTCC-----GCGACCGCGTGGATTCGCGCGCGCTCG 610
QY 325 AspPheGlyTleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAla 344
Db 609 GAC---GGCAGACCGCGCGCGACCTACCGCTGGAATGCGCGCAGCGCGCGCGCGCG 553
QY 345 IleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
Db 552 ATCTCTTCGACATCGCGCAAG-----ACCACCGAGTAC 520
QY 365 SerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGln 394
Db 519 TCCACCGCGTGTGCGCGCAGCAGCAACCGCATC-----CAGGACGCGTACACCATC 472
QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsn 404
Db 471 GGTACGTGCGCGCGCTGGAATCGACGACGCGCGGTGATCTTCGCGCGCTACACCAT 412
QY 405 SerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeu 424
Db 411 GCGCCAGTCCAAAGGTGCGCGCGCGCGTGTGCTGCGCGCAATTCGCCACATCCAGCGCGCTG 352

QY 425 ArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGly 444
Db 351 ACGCGCATCGGCAAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 292
QY 445 LeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnVal 464
Db 291 CGCGCGCTCGGCGACCTCGGCGCTTGCATCCGCGCGCTGAGAGCTCAACGTG 232
QY 465 AspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSer 484
Db 231 GACATCTCAACGAACTGGTGAACCTCATCGTCCACCAGCGCACTACCAAGGCAATGCC 172
QY 485 LysSerValThrAlaAspThrMetLeuGluLysAlaLeuGluLeuLys 501
Db 171 AAGCATTCAGACGAGGATGCGGTGACCCAGACCATCATCAACCTGCGC 121

RESULT 3

US-09-543-681A-4010
; Sequence 4010, Application US/09543681A
; Patent No. 6605709

GENERAL INFORMATION:

; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4010
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-4010

Alignment Scores:

Pred. No.: 1,95e-41 Length: 1239
Score: 475.50 Matches: 136
Percent Similarity: 42.63% Conservative: 81
Best Local Similarity: 26.72% Mismatches: 165
Query Match: 18.39% Indels: 127
DB: 4 Gaps: 15

US-10-009-823A-1 (1-502) x US-09-543-681A-4010 (1-1239)

QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db 40 AGTGGTTTAAATGACGAGCGCGCTAATTAGATACTATTGGTAATAATATCTCAAACTCT 99
QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 100 GCACTTACGTTTAAAGGGGCGACAGTCTCTTTCCGATGTTTGGC----- 150
QY 50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 151 -----GGCTCT-----GGTGCAGGCGTGTGTGTAAAGTATCA 183
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 184 GGCATTAGCCAAACTTTAAAGATGGTAGTATCACTACTACTACCGCCCAACCGATGTG 243
QY 90 AlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLys-----ValHis 106
Db 244 CGCATTTCTGCTGCTGTTTCTTCCTGATT-----GAAGATAGTAAACGCGTGGCGTTTC 297
QY 107 TyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGly 126
Db 298 TATTACGTAACGCTGAATTGGCAAGATAAAACGGCTATCTCAATAATATGCAAGGC 357
QY 127 PheThrLeuMetGlySerArgIleSerAsn-----AsnProhenIleLysLys 142
Db 127 PheThrLeuMetGlySerArgIleSerAsn-----AsnProhenIleLysLys 142

Db 358 ATGCGTATTACAGTTATCCAGTACAAACGAGGTAAATAATGTAGTACAAAGGG 417
QY 143 GluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAla 162
Db 418 GCAACACCAACGCTCATCTATT-----CCTACCGATATGATGAAT---GCA 462
QY 163 LysThrSerThrAlaLeuAsnAlaValValAsnLeuGly-----AspSerThrAsp 179
Db 463 ATGCAACGCTAAATATGATATGATGATGATGATGATGATGATGATGATGATGAT 522
QY 180 LysThrGlnSerGluAlaAsnProTyrPheAlaLeuGluSerTyrLysGlyAsnGly 199
Db 523 CAACACACATAAATTGATCCC-----AAGATATAC--- 555
QY 200 ThrProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGln 219
Db 556 -----GATTCCTATACTTTAGTACTAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 597
QY 220 GlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySer 239
Db 598 TTAGTATATGATATGATATGATATGATATGATATGATATGATATGATATGATAT 639
QY 240 LysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGly 259
Db 640 AAAGATAACGATGAGTGTATGCTCAACAGCACCAACAGGTGAGCCAGCACAGAT 699
QY 260 ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGly----- 277
Db 700 CTG-----GGTAAATAGTTTATAGGATAACGGGGTATTA 735
QY 278 -----GluLeuLeuAsnMetThrAlaPheThrProThrGlySerAlaThr 292
Db 736 GATGAACACGACCTAAGCTGAAACCTTTACCTGCTGCTTATAAGGTTCA----- 789
QY 293 LysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAla 312
Db 789 ----- 789
QY 313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332
Db 790 -----CAACCATGATATGGAATGATACATT----- 816
QY 333 AsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352
Db 816 ----- 816
QY 353 ProSerMetProIleGlnThrSerSerGlyAsnSerThrAlaAlaAsnGlySerSer 372
Db 817 -----AGTGGTAGTACGACGACGACGACGACGACGACGACGACGACGACGAC 849
QY 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392
Db 850 AGTGTCTCTAAACCTCGCGCAAAATGTTATCAACAGCAGGGAATTTACCAATTTCCGTATT 909
QY 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412
Db 910 GAGCCAGATGTTCCATATGCGGACCTACTCAACCAACCAACCAACCAACCAACCAAC 969
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyr 432
Db 970 ATCGCATTAGCTAATTTGCGCAATCTCGTGGCTTAAGTTCAAGGTGACATATGTTGG 1029
QY 433 SerAlaThrLeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 1030 TCTGAACCAATGGTTCAGGCTCACCATTGTTGGTGTGCGAGGCTCTCGCGTATTCGCT 1089
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 1090 AAATTAACCAACAATGCTTTAGAGCTCTAAGGTAGATATGAGCCAGAGGTAGTTAAC 1149
QY 473 MetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 1150 ATGATTGTTGCTCAACGTAATCAATCAATTAATGCGCAGACTCAAAACACTCAGGATCAG 1209

QY 493 MetLeuGlnLysAlaLeuGluLeuLys 501
Db 1210 ATCTGCAACACACTAGTTAGCATGCGC 1236

RESULT 4

US-09-790-988-1
Sequence 1, Application US/09790988
Patent No. 6632935
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: MATANABE, HIDEKI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09790,988
CURRENT FILING DATE: 2003-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Alignment Scores:
Pred. No.: 2,038-30 Length: 640681
Score: 414.50 Matches: 132
Percent Similarity: 39.96% Conservative: 73
Best Local Similarity: 25.73% Mismatches: 187
Query Match: 16.03% Indels: 121
DB: 4 Gaps: 13

US-10-009-823A-1 (1-502) x US-09-790-988-1 (1-640681)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 372153 TCAATATGATAGTATAGTGGCTTACGCAATAATAGTATACATGGAAATTTATATCC 372212
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValAlaPheGlnAsp 43
Db 372213 AATAATATGCTAAGCATCAACTATAGGATATAAATCTCGTAAACCTCTTTTGTAT 372272
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db 372273 ATGTTTCTCATTCATTTTATCAATTAATCAATTAATCAATTAATCAATTAATCAAT 372311
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 372312 TACGAGTGGGTATTTCAAGTATTTATACAAACTTTTAAATGCGCATGTTAGTTGAACT 372371
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGln---ValThrLeuGlu 102
Db 372372 GGCAGCATTTGGATTTAGGAATTTATAAAGACGGCTTTTTCGCTCTTTAGACAGTCAA 372431
QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122
Db 372432 GGTCTATCTCTATATACAGATGGCAATTTCTTCGATTAAGATCAAAATATTATC 372491
QY 123 AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn 136
Db 372492 AATATCAAGTATGATATCAACTGAGCTTATACATCTTTGTTCAAAAGATGATTTTAAAT 372551
QY 137 AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThr 156
Db 372552 AAT-----AGATCAACTTAGAACCTTATTAAATTA-----AAAAATCTTAAT 372593
QY 157 ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly--- 175
Db 372594 ATTTTAAAAAACCAACCTACTTCTGAAATCATCTTAAAGCGTTTTTTTGAATCGTAATACT 372653

QY 176 -----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe 189
Db 372654 GAATCAAAAGCAGGTGGTGAATCTTCACAACTATCTAACACGAGATTACATG 372713
QY 190 AlaLeuLeuGluSerTyrLysGlyAsnGlyThrProProlleSerThrSerAsnTyrSer 209
Db 372714 ACTTATATAGC----- 372725
QY 210 TyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyr 229
Db 372726 -----ATATATAATAAGAGGGAAGGAAAGAGATATATCTGTTCT 372767
QY 230 PheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsn 249
Db 372768 TTTAAT-----AAAAAGAAACAATAAATGACAGTAAAT-----GTGGAATCAAAAT 372815
QY 250 ProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGly 269
Db 372816 GATTCTGATGATAAGACACTATAAAAT-----AGTTTC 372851
QY 270 ThrMetThrPheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly 289
Db 372852 GATTTAACGTTAATGATGATGCGCAATTAACCTCTGATTAATGTTTAAATATATACATCT 372911
QY 290 SerAlaThrLysAspLeuAsnAlaTyrGlnProAlaProLeuValAsnGlyLeuProGln 309
Db 372912 AAAGATTCTAAAG----- 372926
QY 310 PheSerAlaAsnPheValGlyAlaGlyLeuGlnProLeuThrLeuAspPheGlyIleLys 329
Db 372927 -----TATGAAATATCACTTTAAATTTA----- 372950
QY 330 SerGlnGlnAsnMetTyrAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIle 349
Db 372950 ----- 372950
QY 350 GlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsn 369
Db 372951 -----ACAGTACTATAGAACATCAAT 372974
QY 370 GlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAsp 389
Db 372975 TCTGATGTTCTTGGGAAGACATCTCAAAACGATACCTCCAGTAAATTTAAACA 373034
QY 390 ValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAsp 409
Db 373035 TTTGATATTTGTTACTAATGTTGTAATTTGGAACATATTGCAATCAAAACAAACA 373094
QY 410 PheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsn 429
Db 373095 ATAGTCAATATTTATATCAAAATTTATCAATCCAGAAATTTACACCTGAAAGTGT 373154
QY 430 AsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeu---ProGlyThr 448
Db 373155 AATTTATGCTGCTGCTACTGCAAGATCAGGTCAAGCAAAACAGCAATGAAAGCGGTATT 373214
QY 449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArg 468
Db 373215 CAAGATCAGAGTGTAAAGCAATAAAGCGTAGAGTCAAAATGTTGATTTGAATAAA 373274
QY 469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Db 373275 GAAATTAATCAATATGATATATAGCAACAGTAATTTATCAATCTAACGCTCAATCTTTTAA 373334
QY 489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 373335 ACAGNAGATAAATAATTTATACATTAATAATTTACAG 373373

RESULT 5

US-09-252-991A-3144

; Sequence 3144, Application US/09252991A

; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 3144
LENGTH: 810
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3144

Alignment Scores:

Pred. No.: 1,45e-23 Length: 810
Score: 307.00 Matches: 108
Percent Similarity: 30.74% Conservative: 46
Best Local Similarity: 21.56% Mismatches: 93
Query Match: 11.87% Indels: 254
DB: 4 Gaps: 11

US-10-009-823A-1 (1-502) x US-09-252-991A-3144 (1-810)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 25 ATGTATCGGCACCTGTGGTCAGCAACAGCGGTCTGTCGCCCGCAGGACATGAACCTGACC 84
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 85 ACCATTTCACAACTGGCCACGATATCCACACCGGCTTCAAGCGCGACCGCGGAG 144
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 145 TTCAGGACCTGTGTATCCAGATCCGCGCGCCGCGCGCGGACGATGACCCAGGACAGC 204
QY 61 GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla 79
Db 205 GAGTGCTTCGGGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 264
QY 80 PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLys 94
Db 265 TTCACCGCGGCGCAGCTGTCAGACACCGCAGCAGCGCTGACATGCGGTCAACGCGCGC 324
QY 95 GlyPhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPhe 113
Db 325 GCTTCTTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 384
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 385 CACCTGAACCTCCGCGCGCAGATCGTCACCTCCACGCGCTTCGCCCTG----- 432
QY 134 IleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 432 ----- 432
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 433 GAGCCAGCGATC----- 444
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193
Db 444 ----- 444
QY 194 SerTyrLysGlyAsnGlyThrProProlleSerThrSerAsnTyrSerTyrAlaGlnPro 213
Db 444 ----- 444
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233

```

Db 444 ----- 444
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 445 ----- 445
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 460 ----- 460
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 472 ACC----- 474
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 474 ----- 474
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnAsn 333
Db 475 ---GTCGCG----- 480
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 480 ----- 480
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 480 ----- 480
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 481 ----- 481
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 505 ACCACCGGCAACGCCAG-----CCGAGGTGATCGCG-----AACATC 543
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 544 CAGACCGCGCATTCATCAACCCCGCGCGCTCGAGGCCATCGCAACACCTGTTCTCTG 603
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 604 GAACCGGCTCCACCGCGCGCGCTGAGTGGTACGCGCGGTCTCAACGGCCCTCGGCACG 663
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 664 GTTCCCAAGAACACCTCGGAAACTCCACGTCACGTCGTCGAGGAACCTGTTGAACATG 723
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 724 ATACCAACCGCGCGCTACGAGATGAACTCAAGGTCTATCTCCACCGCGCCAGCATG 783
QY 494 Leu 494
Db 784 TTG 786

```

RESULT 6

```

US-09-252-991A-2823/c
; Sequence 2823, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2823
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2823

```

```

Alignment Scores:
Pred. No.: 1.44e-22 Length: 1545
Score: 302.00 Matches: 107
Percent Similarity: 30.60% Conservative: 46
Best Local Similarity: 21.40% Mismatches: 93
Query Match: 11.68% Indels: 254
DB: Gaps: 11

US-10-009-823A-1 (1-502) x US-09-252-991A-2823 (1-1545)

QY 2 MetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThr 21
Db 1545 TTATCGGCACTGTGGTCTGAGCAAGACCGGTCTGTCCGCCAGGACATGAACCTGACCACC 1486
QY 22 ValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPhe 41
Db 1485 ATTTCACAAACCTGCGCAACGTATCCACACCGCGCTTCAAGCGCGAGGAGTTTC 1426
QY 42 GlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGln 61
Db 1425 CAGGACCTGCTGTACCATCGCGCGCCAGCGCGCGCCAGTGCACCCAGGACAGCGAG 1366
QY 62 AlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAlaPhe 80
Db 1365 CTGCTCTCGGCGCTGCAACTGGTACTCGGTGCGGTGCGTGGCGCACCCAGAGATCTTC 1306
QY 81 GluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLysGly 95
Db 1305 ACCCGCGGCGCTGCGACACCGAGCGCGCTGGACATGGCGGTCAACGGCGCGCGGC 1246
QY 96 PhePheGlnValThrLeuGluAsp---LysValHisTyrThrArgAlaGlyAsnPheArg 114
Db 1245 TTCTTCCAGGTCTGCTCGCGGACCGACCGTGTCTTACACCGCGCGAGCGGAGCTTCCAC 1186
QY 115 PheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArgIle 134
Db 1185 CTGAACCTCGGCGCGGAGATCGTCACTCCACAGCTTCGCCCTG----- 1141
QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
Db 1140 -----GAG 1138
QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
Db 1137 CCACCGCATC----- 1129
QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
Db 1129 ----- 1129
QY 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
Db 1129 ----- 1129
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
Db 1129 ----- 1129
QY 235 SerSerThrGlySerLysThrPheGluTyrLeuValaMetAsnProSerGluAspGly 254
Db 1128 -----GTGGTGGCCCAACGAG----- 1114
QY 255 SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
Db 1113 -----ACCCAGACCTCTCACC 1099

```

QY 275 SerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAsp 294
Db 1099 ----- 1099
QY 295 LeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPhe 314
Db 1099 ----- 1099
QY 315 ValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMet 334
Db 1098 GTCCGC----- 1093
QY 335 TrpAlaGlyAlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSer 354
Db 1093 ----- 1093
QY 355 MetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThr 374
Db 1093 ----- 1093
QY 375 ArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSer 394
Db 1092 -----CAGGACGCG-----ACCCTCTCGGTGACCC 1066
QY 395 GluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePro 414
Db 1065 ACCGCCAACCCCGAG-----CCGAGGTGATCGC-----AACATCCAG 1027
QY 415 LeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAla 434
Db 1026 ACCGCCGACTTCATCAACCCGCGCGCGCTCGAGCCCATCGCAACACCTGTTCTCTGGAA 967
QY 435 ThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeu 454
Db 966 ACCGGCTCCAGCGCGCGCGCGAGGTGCGTACCGCGGTCTCAACGGCCTCGGACGGTT 907
QY 455 SerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIle 474
Db 906 GCCCAGAACACCTTGGAAACTCCACGCTCAACGTGCTGAGGAAGTGTGTAACATGATC 847
QY 475 IleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu 494
Db 846 ACCACCCAGCGCGCTACGAGATGAATCCAAAGTCACTCCACCGCCGACGATGTTG 787

RESULT 7

US-09-543-681A-3938
; Sequence 3938, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3938
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-3938

Alignment Scores:
Pred. No.: 2,6e-21 Length: 798
Score: 286.00 Matches: 100
Percent Similarity: 30.57% Conservative: 55
Best Local Similarity: 19.72% Mismatches: 98
Query Match: 11.06% Indels: 254
DB: 4 Gaps: 9

US-10-009-823A-1 (1-502) x US-09-543-681A-3938 (1-798)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 16 ATGATCCGCTCTTATGATTCGTAACACAGGTTGGATGCACCAACCAACTAAACATGAT 75
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 76 GTGATTCACCAACACTCGCAACGTCAGCAACAATGGTTTAAACGCCAGCGTGGGTT 135
QY 41 PheGlnAspLeuPheSerGlnAspLeu-----AlaIleGlySerThrGlySer 56
Db 136 TTTGAGGATTTACTCTATCAAACTATTTCGTCACCGGAGCGATGACATCCGACGACG 195
QY 57 GlnGlyProAsn-----GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
Db 196 AATGCGCTTCTGTTTACAAATGGTACTGCTGCTGCTGCCAGTGGCAGACAGCATT 255
QY 75 PheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys 94
Db 256 CATAGCCCAAGTAATTTAGCCCAAACTAATGGTACGGTGATGTTGCTATCAAGGGCAA 315
QY 95 GlyPhePheGlnValThrLeuGluAspLysValHis---TyrThrArgAlaGlyAsnPhe 113
Db 316 GGTITTTTCCATGTTCAATTACCTGATGGTACGATGCTTATATCTCGTATGGCTCTTT 375
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 376 CAAATGGACCAAAATGGSCAACTAGTGACCTCCAGTGGCTTTCAAATCGTCCAGCG--- 432
QY 134 IleSerAsnAsnProAsnIleLysGlnThrLeuGluProIleGlnLeuAspPheAsn 153
Db 433 -----ATTATTTTGCAGAAACCGCTAAAGAGGTGATGGTA----- 468
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 468 ----- 468
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193
Db 468 ----- 468
QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
Db 468 ----- 468
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
Db 469 -----GGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 507
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 508 CTGCA----- 513
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 513 ----- 513
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 513 ----- 513
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 514 -----CCTCA----- 519
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333
Db 519 ----- 519
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 520 -----CAAGTAGGGCAATTA--- 534

QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 534 ----- 534
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 535 -----ACCTTCACC 543
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 544 ACA----- 546
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 547 -----TTTATTATGATAGCGGGTTAGAAAGTGTGGGAAATCTGTACTTA 594
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 595 GAAACAGCCAGCTCCGAGACCCCACTGAGAAATCGCGGGTATTAAACGGTGGCGGCTG 654
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 655 TTATATCAGGATATGTTGAACCTCTAACCGTTAATGTCGCGAAGAAATGGTCAATATG 714
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493
Db 715 ATCCAACTCAGCGTGTATTGAATTAATTAAGTAAAGGATTCCACTTCTGATCAGATG 774
QY 494 LeuGlnLysAlaLeuGluLeu 500
Db 775 TTACAGAACTCAGCGCACTC 795

RESULT 8

US-08-483-857-1
; Sequence 1, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPILOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:

NAME/KEY: CDS
LOCATION: Join(101..910, 942..1730)
US-08-483-857-1

Alignment Scores:
Pred. No.: 3 01e-20 Length: 1800
Score: 281.50 Matches: 133
Percent Similarity: 38.41% Conservative: 94
Best Local Similarity: 22.50% Mismatches: 228
Query Match: 10.89% Indels: 136
DB: 3 Gaps: 22

US-10-009-823A-1 (1-502) x US-08-483-857-1 (1-1800)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 101 ATGCAAAATGGATATATATCAAGCAACTCGCGGAATGGTAACCTGATTTAATAAACTTGAT 160
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 161 GTGATTACTAATAATCTTGCCAAATATCAATCACTGATATATAAGAGATGATGTGTT 220
QY 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlaIleGlySer--- 53
Db 221 ATTCAGATTTTAAAGGATTTTAAAGAACTCAGGATGAGTTGCCCTATAGAAAATCAC 280
QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly---AlaGlnValGlySer 70
Db 281 ACAAGAGATGATCTCGTTTGTAAATCTACATAGATGGAATCCCAAGTTCTCTCAA 340
QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
Db 341 GAAATATACGATTTTAGCTAGTTCTTTAAAGGCGCACAAACAATCTCTTGGATTGGCA 400
QY 91 IleGlyGlyLysGlyPhePheGlnVal-----ThrLeuGluAspLysValHisThr 108
Db 401 ATGACTAGAGAGATGCTTTTATTTGTTTCAGACCAAGATGAGAGATGAAGATTAAAC 460
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 461 AAAGATGGAATTTTCAACTTGATGATGAGGTTATTGGTAAATAAGCAAGGATACAAG 520
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsn----- 139
Db 521 GTATTAAAGTAGTATGATTTTAAATAATCTCAGAAATGCTGCATACCATCTCTTAATAGT 580
QY 139 ----- 139
Db 581 GCTGTTCAATTAGCGTTGATAAAAACGGAAGCATTGAAGTTGATGGAGTCAAAATGCA 640
QY 140 -----IleLysLysGluThrLeuGluProIleGlnLeuAspPheAsn----- 153
Db 641 AGATTATTGTTAGCACACAGATGATATAGAGCTTTGCAAAAAGATGGGGATAATGTC 700
QY 154 -----AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAla 170
Db 701 TATAAATAGATGATCTAACCCGTTATTAGAGATTGAAAACCTCAATGCTATTTCGCCAA 760
QY 171 ValValAsnLeuGlyAspSerThrAspLysThrGlnSer-----GluAla 185
Db 761 GGTTTTCTCAGGGATCAATGTTAATCCAGTTACTGAAATGGTAGGACTGATTGAAGCA 820
QY 186 AsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProProIleSerThr 205
Db 821 AAC-----AGAAATGGTAGAAATGTATCAAAAAGTTATGACAGCTCATATGGATGAC 871
QY 206 SerAsnTyrSerTyrAlaGlnProMet----- 214
Db 872 TTAATCAAGAAGCTATCAATAAGCTGTAAATAATTTAAATAATAATAAAAA 931
QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
Db 932 AGGATTTAAATATGATGATCATCTTCATAC-----TGCTGCTACAGGAATGGTAGCGCA 985

Db 1 ATGATGAGATCACTTCTATCTAGTCTGCTAGCAATGCTAGCGCAGCAACAAATTCAT 60
QY 21 ThrValSerAsnIleAlaAsnAlaAsnThrIleGlyTyrIleGlnGlnValVal 40
Db 61 GTTACTTCAATCAATCAATGCGCAATGTTAATACAGCAGGTTTAAAGAAAGTCCGCGA 120
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 121 TTTGCTGATCTTATGATCAAGTTATGAAGTATGCAGGAACCTCAACTTCAGCTACTACT 180
QY 61 GlnAlaGlyMetGlyAlaGlnValGly 74
Db 181 CTTTCTCCTTCCTGCGGTATAGAGTGGGTGGGTGGTGGTCCACAGCGGTAACTAAAGTT 240
QY 75 PheThrGlnGlyAlaPheGlu 200
Db 241 TTTTCTGAAGAAATTTAAATCAACAGTACTGATGGTCTTGTATGCTATTCAGCT 300
QY 94 LysGlyPheGlnValThrLeuGluAsp 300
Db 301 AATGGGTGTTTTCAAATACAACTTCCTGATGGCACTATAGGATATAGGATATAGGCA 360
QY 113 PheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySer 132
Db 361 TTTACAAAGATATAGGATATATGTAATTCAGATGCTATAGACTT 411
QY 133 ArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPhe 152
Db 411 411
QY 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal 172
Db 411 411
QY 173 AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeu 192
Db 411 411
QY 193 GluSerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGln 212
Db 411 411
QY 213 ProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGly 232
Db 411 411
QY 233 AlaProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGlu 252
Db 411 411
QY 253 AspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThr 272
Db 411 411
QY 273 PheSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThr 292
Db 411 411
QY 293 LysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAla 312
Db 412 412
QY 313 AsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGln 332
Db 421 421
QY 333 AsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu 352
Db 439 439
QY 353 ProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 372

Db 475 TCTGTAAATGCTACCA 489
QY 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392
Db 489 489
QY 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412
Db 490 490
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyr 432
Db 520 GTGGAGCTAGTTCAGTTTATAATCCAGCGGCTCTTATCTATCGGTGATAATCTTTAT 579
QY 433 SerAlaThrLeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 580 CTTGAACAGGAGCAAGTGGTCCACCTGTTCCGGGTATAGCAGCAGATGGGCTTGA 639
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 640 ACAATAAGACATGGATTATAGAACTTAGTAATGTTCAAGCTTGTGAAGAAATGACAGAT 699
QY 473 MetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 700 CTTATCACAGGACAAAGAGCTTATCAAGCGGTTTCAAGCGCTTATCAACCAAGTGATGAT 759
QY 493 MetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 760 ATGCTAGGAATGTAAATCAGCTTAAGCGA 789

RESULT 11
US-08-436-748-4
; Sequence 4, Application US/08436748
; Patent No. 5827654
; GENERAL INFORMATION:
; APPLICANT: CHAN, YOON LOONG
; APPLICANT: LOUIE, HELENA
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,748
; FILING DATE: 05-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-428 MIS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-436-748-4

Alignment Scores: 1.4e-14 Length: 789
Pred. No.:

Score: 223.50 Matches: 98
Percent Similarity: 27.70% Conservative: 43
Best Local Similarity: 19.23% Mismatches: 115
Query Match: 8.64% Indels: 253
DB: 1 Gaps: 12

US-10-009-823A-1 (1-502) x US-08-436-748-4 (1-789)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 1 ATGATGAGATCACTTCATCTCTCTAGCAATGGTAGCGCAACACAAATGAT 60
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 61 GTTACTTCAATAACATCGCAATGTAATACAGCAGGTTTAAAGAAAGTCGCGCAGAA 120
QY 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 121 TTGTCTGATCTATGATATCAAGTTATGAAGTATGATGAGCAACTTCAACTTCAGCTACTACT 180
QY 61 GlnAlaGlyMetGlyAlaGlnValGly-----SerValArgThrIle 74
Db 181 CTTTCTCCTCGGTATAGAGTGGTGTGGTGTGCTGCCAACAGCGGTAAAGTT 240
QY 75 PheThrGlnGlyAlaPheGlu---ProGlyAsnSerValThrAspLeuAlaIleGlyGly 93
Db 241 TTTACTGAAGGAATTTAAATCAACAAAGTACTGATGCTTGTATGCTTATGCTTATGCGAGGT 300
QY 94 LysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPhe 113
Db 301 AATGGGTTTTTCAATCAACTT----- 324
QY 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 325 -----CCTGATGCGCACTATAGAAATGGGC----- 348
QY 134 IleSerAsnProAsnIleLysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 349 -----AATTTACAAAG----- 360
QY 154 AspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsn 173
Db 361 -----ATAATCAAG----- 369
QY 174 LeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGlu 193
Db 369 ----- 369
QY 194 SerTrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnPro 213
Db 369 ----- 369
QY 214 MetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAla 233
Db 370 -----GATAATCAAGTAATATT----- 387
QY 234 ProSerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAsp 253
Db 387 ----- 387
QY 254 GlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 388 -----GTAATTCAGATGGTTATAGACTTTTACCTGAATGACAATA 429
QY 274 SerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys 293
Db 430 -----CCTGAAGGC----- 438
QY 294 AspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsn 313
Db 438 ----- 438
QY 314 PheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsn 333

Db 438 ----- 438
QY 334 MetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuPro 353
Db 439 -----GCAACAGCAATTAATGTGTCTACAGAT---GGAACCGTTTCT 477
QY 354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSer 373
Db 478 GTAATGCTACCA----- 489
QY 374 ThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThr 393
Db 489 ----- 489
QY 394 SerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIle 413
Db 490 -----GGGAGCAACAAGAACTCAATT-----GGCCAAGTG 522
QY 414 ProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSer 433
Db 523 GAGTAGTTCAGTTTATAATCCAGCGGTCTTTCATTCTATGGGTGATAATTTATCTT 582
QY 434 AlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db 583 GAAACAGGAGCAAGTGGTGCACCTGTTGCGGTATAGCAGGACAGATGGGCTTGGAAACA 642
QY 454 LeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMet 473
Db 643 ATAAGACATGGATTATAGACTTAGTAATGTTTCAGCTTGTGAGAAATGACAGATCTT 702
QY 474 IleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMet 493
Db 703 ATCAGACAGCAAGAGCTTATGAGCGGTCTTCAAGCGCATTACCAACAAAGTGATGATG 762
QY 494 LeuGlnLysAlaLeuGluLeuLysArg 502
Db 763 CTAGGAATGTAAATCAGCTTAAGCGA 789

RESULT 12.
US-09-689-065B-2
; Sequence 2, Application US/09689065B
; Patent No. 6605696
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products, Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS AND RELATED METHODS AND
; TITLE OF INVENTION: MATERIALS
; FILE REFERENCE: 3153.00187/PC10589A
; CURRENT APPLICATION NUMBER: US/09/689,065B
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US Prov. 60/160,922
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US Prov. 60/163,858
; PRIOR FILING DATE: 1999-11-05
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 5445
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-09-689-065B-2

Alignment Scores:
Pred. No.: 6,45e-10 Length: 5445
Score: 193.00 Matches: 40
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.46% Indels: 0
DB: 4 Gaps: 0

US-10-009-823A-1 (1-502) x US-09-689-065B-2 (1-5445)

QY 463 AsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMet 482

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 3047
LENGTH: 408
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3047

Alignment Scores:
Pred. No.: 1,04e-09 Length: 408
Score: 174.00 Matches: 43
Percent Similarity: 50.00% Conservative: 27
Best Local Similarity: 30.71% Mismatches: 64
Query Match: 6.73% Indels: 6
Dbs: 4 Gaps: 3

US-10-009-823A-1 (1-502) x US-09-252-991A-3047 (1-408)

Qy 34 TyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySer 53
Db 3 TTCAGCAATCCCGCGGAGTTCCGCGACGCTACGCCGCCCTCGGTG---CTGGGTTCG 59
Qy 54 ThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThr 73
Db 60 -----GGCAGCAACCCCGAGCGGCGGCGTCTCGGACGTCCTCGCAG 107
Qy 74 IlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGly 93
Db 108 ATGTTCAAGCAGGCAATCATGCTCCACCAACAGCGTGTGGACCTGGCCATCAAGGC 167
Qy 94 LysGlyPheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPhe 113
Db 168 AACGGCTTCTTCGTCACCAACCAACAGGGGCGGATCAGCTACACCCCGCGGCTACTTC 227
Qy 114 ArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMetGlySerArg 133
Db 228 AATACCGCAACAGCAGGATTTCATGCTGACAAACACCGCTACCGCTCGAGGGGTATGCC 287
Qy 134 IleSerAsnAsnProAsnIleLysGlnThrLeuGluProIleGlnLeuAsp---Phe 152
Db 288 CTCGGCGCAACGCCAGTTCAGAACCGCGTGTCCACCGACCTCAAGGTCGAGCGGCC 347
Qy 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaVal 172
Db 348 AATCAGGCGCGCAGGCGCACCTCGAGCATCCAGCGTGTACAAACCTCAACTCGCGCTG 407

RESULT 16
US-09-206-942-68
Sequence 68, Application US/09206942
Patent No. 6432669
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
TITLE OF INVENTION: Molecular Weight Proteins
FILE REFERENCE: 1038-861 MIS:jb
CURRENT APPLICATION NUMBER: US/09/206,942
CURRENT FILING DATE: 1998-12-08
EARLIER APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 68

LENGTH: 3285
TYPE: DNA
ORGANISM: Haemophilus influenzae
US-09-206-942-68

Alignment Scores:
Pred. No.: 3,58e-08 Length: 3285
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
Dbs: 4 Gaps: 24

US-10-009-823A-1 (1-502) x US-09-206-942-68 (1-3285)

Qy 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49
Db 1807 AATATTTTCAGGTTTCATATAAAGCAGATTACAGTTAAAGAT-----GGTAGTGATTTA 1860
Qy 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
Db 1861 ACTATTGGTAACCAATAGTGTCTGATGTTACTAAT----- 1896
Qy 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
Db 1897 ---GCCAAAAAAGTAACCTTT-----AACCCAGGTTAAAGAT 1929
Qy 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
Db 1930 TCAAAAATCTCTGCTGACCGGT---CACAAAGTGACACTACACAGCAAGTAGTGA 1980
Qy 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 1981 ACATCCGGTAGTAATAACACACACTAGATAGCAGTACAATAAT---GCCGGCTTA 2037
Qy 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlnThrLeuGluProIle 148
Db 2038 ATCGATGCAAAAATGTAACAGTAAACCAACATATTACTTCTCAACAGCAGTGAGCATC 2097
Qy 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 2098 -----TCTCGCAAGTAGTGAGAAATACCCTAAACAGGTA 2142
Qy 169 AsnAlaValAlaAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 2143 AACGCCAACCTGTTAACTGGAGATAACCGCTCAACAGGTA 2190
Qy 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203
Db 2191 CTAGTGGGAATTGAGTCCAGCTCTGCTCTGTAACACTTACTGCAACCGCGGCTCTT 2250
Qy 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 2251 GCTGTAAAGCAATATTTCG-----GCCAACACC 2277
Qy 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 2278 GTTACTGTTACTGCAAT---AGCGGTGCAATTAACCACTTTCGCGAGGCTCTCAATAA 2334
Qy 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 2335 GGAACCGAGGTGTACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTCTTCTGT 2394
Qy 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
Db 2395 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTAAACCACTCAATCCCAATTCAAAAATTAA 2454
Qy 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
Db 2455 GCAACACAGCGGCGAGGCTACGTAAACAGTGCACAGGTACATTCGTTGGTACGATTCC 2514
Qy 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320

Db 2515 GGTAAATACGTAAT-----GTTACGGCAAAACGCTGGC 2547
QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
Db 2548 GATTAAACAGTTGGGATGGCGAGAA-----ATTAAATGGAGCAAA 2589
QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db 2590 GGAGCTGCAACCTTAACATCATCATCGGCAATTA-----ACT 2628
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp 380
Db 2629 ACCGAGCTAGTTACACATTAACCTTACGCAAGGTCAGTAAATCTTTTCAGCTCAGGAT 2688
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 2689 GGTAGCGTTGCGAGGAATTAATGCGCCAAATGTGACACTAAATACTACAGGCACITTA 2748
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 2749 ACTACCGTGAAGGGTTCAACACATTAATGCAACCGGTACCTTGGTTATTATACGCAAAA 2808
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
Db 2809 GACGCTGAGCTAAATGGCGCAGCATTTGGTAAACACACAGTGGTAAATGCAACCAACGCA 2868
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Db 2869 AATGCTCCGCGCAGCGTAAATCGCGACACCTCAAGC-----AGAGTGAACATCACT 2919
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
Db 2920 GGG-----GATTAAATCAATAATGGAITTA-----AATATC-----ATT 2955
QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetMetSerLysSer 486
Db 2956 TCMAAAACCGGTATAAACACCGTACTGTTA---AAAGCGGTAAATGATGTGAATATAC 3012
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 3013 ATTCAACCGGTATAGCAAGCGTAGATGAATGAATTAATTGAAGCAAAACGC 3060

RESULT 17

US-08-038-682-1
; Sequence 1, Application US/08038682
; Patent No. 5549857
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-038-682-1

Alignment Scores:

Pred. No.: 7.39e-08 Length: 5116
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 1 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-038-682-1 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 3480 AATATTTTCAGGTTTCATTAACACAGAGATTACAGCTTAAGAT-----GGTAGTGATTTA 3533
QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
Db 3534 ACTATTTGGTAACCAATAGTGTCTGATGCTACTAAT----- 3569
QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
Db 3570 ---GCCAAAAAAGTAACCTTT-----AACAGGTAAAGAT 3602
QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108
Db 3603 TCMAAAATCTCTGCTGACGGT---CACAGGTGACACTACACAGCAAAAGTGAA----- 3653
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 3654 ACATCCGGTAGTAATAACACACACTCAAGATAGCAGTGAATAAT---GCCGGCTTAAT 3710
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
Db 3711 ATCGATGCAAAAAAATGTAACAGTAAACACATATTACTTCTCACAAGCAGTGAATC 3770
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 3771 -----TCTCGCAGCAAGTGGAGAAATTAACACTAAACAGGTACACCAAT 3815
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 3816 AACGCAACCACTGGTAACGTGGAGATACCTCAACAGGTAGT-----ATC 3863
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProIle 203
Db 3864 CTAGGTGGAATTAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGGGCGCTCT 3923
QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 3924 GCTGTAAAGCAATATTTCG-----GGCAACACACC 3950
QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 3951 GTTACTGTACTGCAAAAT---AGCGGTGCAATTAACCACTTTGGCAGGCTCTACAAATAA 4007
QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 4008 GGAACCGAGAGTGAACCACTTCAAGTCAATGACGGCATATCGCGGTAGCATTTCTGGT 4067
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
Db 4068 GGCACAGTAGAGGTAAAGCAACCGAAAGTTTAACTCAATCAATTCAAAATAA 4127

QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 3951 GTTACTGTTACTGCAAT---AGCGGTGATTAACACCTTTGGCAGGCTCTACAATTA 4007
QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
Db 4008 GGNACCGAGGTATACCACTTCACTCACTCACTCACTCACTCACTCACTCACTCACT 4067
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
Db 4068 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTAAACCACTCACTCACTCACTCACTCACT 4127
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
Db 4128 GCAACACAGCGGAGGCTAACGTAAACAGTCAACAGGTACCAATTTGGTACGATTTCC 4187
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
Db 4188 GGTAAACGGTAAAT-----GTTACGCAACCGCTGGC 4220
QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
Db 4221 GATTTAAACAGTTGGGAATCGGCAGAA-----ATTAATGCGACAGAA 4262
QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db 4263 GGAGCTGCAACCTTAACATCATCTCGGCAGAA-----ACT 4301
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgGlySerGlnAsp 380
Db 4302 ACCGAAGTGGTTCACATTAACACCAAGGTCAGGTAATCTTTACGCTCAGAT 4361
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 4362 GGTACGGTTCAGGAAGTATTAAATGCCCAATGTACACTAAATATCTACAGGCACCTTA 4421
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 4422 ACTACCGTGAAGGTTCAACATTAATCAACACCGGTACCTTGGTTATTATTAACGCAAA 4481
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
Db 4482 GACGTGAGCTAAATGCGCGAGCATTTGGTAAACACACAGATGGTAAATCAACCAACGCA 4541
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Db 4542 AATGGCTCCGCGAGCGTAAATCGGACACCTCAAGC-----AGAGTGAACATCACT 4592
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
Db 4593 GGG-----GATTTAATCAACAATAATGATTA-----AATATC---ATT 4628
QY 467 SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
Db 4629 TCAAAAACCGGTATAACACCGTACTGTTA---AAAGCGGTAAATTAATGATGTAATAC 4685
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuLysArg 502
Db 4686 ATTCAACCGGTATAGCAAGCGTATAGTAAATTAATGAAAGCAACGC 4733

RESULT 19

US-08-530-198-1
; Sequence 1, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/530,198

FILING DATE: 13-DEC-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: BERNSTRESSER, JERRY W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: JWB-1186

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5116 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-10-009-823A-1

(1-502) x US-08-530-198-1 (1-5116)

Alignment Scores:

Pred. No.: 7,39e-08 Length: 5116

Score: 173.50 Matches: 110

Percent Similarity: 37.30% Conservative: 75

Best Local Similarity: 22.18% Mismatches: 210

Query Match: 6.71% Indels: 101

DB: 2 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-530-198-1 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49
Db 3480 AATATTTTCAGGTTTCAATAAAGCAGAGATTACAGCTAAAGAT-----GGTAGTGATT 3533
QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
Db 3534 ACTATTGGTAAACCAATAGTCTGCTGATGCTACTAAT----- 3569
QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
Db 3570 ---GCCAAAAAAGTAACCTTT-----AACCAGGTTAAAGAT 3602
QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisThr 108
Db 3603 TCAAAAATCTCTGTCGACGGT---CACAAAGGTGACACTACACAGCAAGTGA--- 3653
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 3654 ACATCCGGTAGTAAATAACACACTGAAGTAGCAGTACATAAT---GCCGGCTTAAT 3710
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysGluThrLeuGluProIle 148
Db 3711 ATCGATGCAAAAATGTAAACAGTAAACCAATATTACTTCTCACAAGCAGTACGATC 3770
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 3771 -----TCGCGCAAGTGGAGAAATTACCCTAAACAGGTACACCAT 3815
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 3816 AACGCAACCACTGGTAAACGTGAGATAACCGCTCAACAGGTAGT-----ATC 3863

189	QY	189	PheAlaLeuLeuGluSerTrpTyrGlyAsnGlyThr-----ProPheLeu 203
3864	Db	3864	CTAGGTGAAATTGAGTCCAGCTCTGCTCTGTAAACACTTACTGCAACAGGCGCTCTTT 3923
204	QY	204	SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnIysAsnSer 223
3924	Db	3924	GCTGTGAAGCAATATTTCC-----GGCAACACC 3950
224	QY	224	HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
3951	Db	3951	GTTACTGTTACTCAAAAT---ACCGGTGATTAAACCACTTTGGCAGGCTCTACAAATTAAA 4007
244	QY	244	TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
4008	Db	4008	GGAACCGAGAGCTGTACCACTTCAAGTCAATCAGCGCATATCGCGGTACGATTTCTGGT 4067
264	QY	264	GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
4068	Db	4068	GCCACAGTAGAGGTTAAAGCAACCGAAAGTTAAACCACTCAATCCAAATTCAAAAATTAAA 4127
281	QY	281	AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
4128	Db	4128	GCACCAACAGCGAGGCTAACGTAACAGTGCACACAGGTACAAATTTGGTGGTACGATTTCC 4187
301	QY	301	AlaProIleuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
4188	Db	4188	GGTAATACGGTAAAT-----GTTACGGCAACACGGTGGC 4220
321	QY	321	ProLeuThrLeuAspPheGlyIleLysSerGlnAsnMetTrpAlaGlyAlaProAla 340
4221	Db	4221	GATTACAGTTGGGATGCGCGAA-----ATTATGCGACAGAA 4262
341	QY	341	SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
4263	Db	4263	GGAGCTTCAACCTTAACATACATCATCGGCAAAATTA-----ACT 4301
361	QY	361	SerSerGlyAsnSerThrAlaAspAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
4302	Db	4302	ACCGAAGTAGTTCACACATTCTTCAGCCAGGTCAGGTAATCTTCAGCTCAGGAT 4361
381	QY	381	GlyTyrProGlnIleAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
4362	Db	4362	GSTAGCGTTCAGAGAAATTAATTCGCGCAATGTGCACACTAAATACTACAGGCACCTTA 4421
399	QY	399	-----GlnGlyLys-----TyrSerAsnSerGln 406
4422	Db	4422	ACTACCGTGAAGGTTCAACATTANTCAACCAGCGGTACTTGGTTATTATACGCAAAA 4481
407	QY	407	ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
4482	Db	4482	CACCGCTAGCTAAATGCGCGAGCATTTGGGTACACACAGTGGTAAATGCAACCAACGCA 4541
427	QY	427	GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
4542	Db	4542	AATGGTTCGGCAGCGTAAATCGCGACCACTCAACG-----AGAGTGAACATCACT 4592
447	QY	447	GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
4593	Db	4593	GGG-----GATTTAATCACAATAATGGATTA-----AATATC---ATT 4628
467	QY	467	SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
4629	Db	4629	TCRAAAAACGGTATTAACACCGTACGTGTTA---AARGCGGTAAATTAATGATGAATAC 4685
487	QY	487	ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
4686	Db	4686	ATTCAACCGGTATAGCAACGGTAGATGAAGTAATTGAAGCAACCG 4733

RESULT 20

RESULT 20
US-08-469-880-1

US-08-469-880-1
; Sequence 1, Application US/08469880

Patent No. 5876733

GENERAL INFORMATION:

1000

[REDACTED]

```

; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MJS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-1

Alignment Scores:
Pred. No.: 7.39e-08 Length: 5116
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: Gaps: 24

US-10-009-823A-1 (1-502) x US-08-469-880-1 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAspLeuPheSer
Db 3480 AATATTTCAGTTCAATAAAGCAGAGATACAGCTAAAGAT-----GGT
QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGly
Db 3534 ACTATTGGTAACACCAATAGTGTGTGATGGTACTAAT
QY 69 GlySerValArgThrIlePheThrGlnGlnValAlaPheGluProGlyAsnSer
Db 3570 ---GCCAAAAGTAACCTTT-----AACCA
QY 89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGlnAspLysVal
Db 3603 TCAAAAATCTCTGCTGACCGT---CACAAGGTGACACTACACAGCAAGT
QY 109 ArgAlaGlyAsnPheArqPheThrGlnAspGlyPheLeuAsnAspProSer

```

RESULT 20

RESULT 20
US-08-469-880-1

US-08-469-880-1
; Sequence 1, Application US/08469880

Patent No. 5876733

GENERAL INFORMATION:

1000

[REDACTED]

```

Db 3654 ACATCCGGTAGTAATAACACACTGAAGATACGACGTGCAATAAT---GCCGGCTTAAC 3710
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
Db 3711 ATCGATGCAAAATGTTACAGTAAACAAATATTAATCTTCCAAAGCAGTGGCAGTC 3770
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
Db 3771 -----TCTGCGCAAGTGGAGAAATTAACCACTAAACACAGGTACACCAATT 3815
QY 169 AsnAlaValAlaLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
Db 3816 AACGCAACCACTGTTAAACGTGGAGATAACCGCTCAACAGGTAGT-----ATC 3863
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203
Db 3864 CTAGTGGAAATGAGTCCAGCTCTGCTCTGTAACACTTACTGCAACCGAGGCGCTCTT 3923
QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
Db 3924 GCTGTAGCAATATTTCG-----GGCAACACCC 3950
QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
Db 3951 GTTACTGTACTGCAAAAT---AGCGGTGCAATTAACCACTTTGGCAGGCTCTACAAATTA 4007
QY 244 TyrLeuValAlaMetAsnProSerSerGluAspGlySerAlaLysSerGlyThrAspSerAla 263
Db 4008 GGAACCGAGAGTAAACCACTTCAAGTCAATCAGCGGATATCGCGGTACCACTTTCTGGT 4067
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
Db 4068 GGCACAGTAGAGTTAAAGCAACCGAAGTTTAAACCACTCAATCAATTCATAAATTA 4127
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
Db 4128 GCAACACAGCGAGGCTTAACGTAACAGTGCACACAGGTACAAATTTGGTGTGATTC 4187
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
Db 4188 GGTAATACGGTAAT-----GTTACGGCAACAGCTGGC 4220
QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
Db 4221 GATTTAAACAGTTGGAAATGGCGCAGAA-----ATTAATGCGACAGAA 4262
QY 341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
Db 4263 GGAGCTGAACCTTAATCACTATCATCGGCAATTA-----ACT 4301
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
Db 4302 ACCGAAGTAGTTCCACACATTACTTCAGCAAGGTCAGGTAAATCTTCAGCTCAGAT 4361
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 4362 GGTAGCTGCGAGGAAGTATTAATGCGCAATGTGACACTAAATACACAGCACTTTA 4421
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 4422 ACTACCGTAGAGGTTCAACATTAAATGCAACCGGTACCTTGTATTATTAACGCAAA 4481
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426
Db 4482 GAGCTGAGCTAATAGCGCAGCAATTTGGGTAAACACACAGTGTGTAATGCAACCAACGCA 4541
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Db 4542 AATGGCTCCGCGCAGCTAATCGGCAACCACTCAAGC-----AGAGTGAACATCACT 4592
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466

```

```

Db 4593 GGG-----GATTTAATCAATAAATGGAATTA-----AATATC---ATT 4628
QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
Db 4629 TCAAAAACGGTATAAACACACCGTACTGTTA---AAAGGGTTAAATTCATGTGAATAC 4685
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 4686 ATTCAACCGGTATAGCAAGCGTAGATGAAGTAATTAAGCGCAACGC 4733
RESULT 21
US-08-728-470-1
; Sequence 1: Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkettreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-728-470-1
Alignment Scores:
Pred. No.: 7,39e-08 Length: 5116
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservative: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 2 Gaps: 24
US-10-009-823A-1 (1-502) x US-08-728-470-1 (1-5116)
QY 30 AsnThrIleGlyTyrLysGlnGlnValPheGlnValPheGlnAspLeuPheSerGlnAspLeu 49

```

3480 ATATTTCAGGTTTCAATTAAGACAGAGATTACAGCTAAGAT-----CGTAGTGATTTA 3533
50 AlaileGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
3534 ACTATTGGTAACCAATAGTGTGATGCTACTAAT----- 3569
69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
3570 ---GCCAAAAAGATAACCTTT-----AACAGGTTAAAGAT 3602
89 LeuAlaIleGlyGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyThr 108
3603 TCAAAATCTCTGCTGACGCT---CACAAAGTGCACACTACACAGCAAAAGTGGA----- 3653
109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
3654 ACATCCGGTAGTAAATACACACTGAGATAGCAGTGACAAATAAT---GCCGCTTAAC 3710
129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
3711 ATCATGTCAAAAATGTACAGTAAACAAATAATTACTTCTCACAAAGCAGGAGCATC 3770
149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
3771 -----TCTGGCACAAGTGGAGAAATTACCCTAAACACAGGTACACACCAATT 3815
169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyThr 188
3816 AACGCAACCACTGTAACGTGGAGATACCGCTCAACACAGTAGT-----ATC 3863
189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203
3864 CTAGGTGGAATTGAGTCCAGCTCTGCTCTGTAACTTACTGCAACCGGCGCTCT 3923
204 SerThrSerAsnTyThrValAlaGlnProMetArgValTyThrAspGlnGlyAsnSer 223
3924 GCTTAAGCAATATTCG-----GGCAACACC 3950
224 HisAspIleThrValTyThrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
3951 GTTACTGTACTGCAAT---AGCGGTGCAATTAACCACTTGGCAGGCTCTACAAATTA 4007
244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaLysThrAspSerAla 263
4008 GGAACCGAGAGTAAACCACTTCAAGTCAATCAGCGCATATCGCGGTACGATTTCGT 4067
264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
4068 GGCACAGTAGAGTTAAGCAACCGAAGTTTACCCTCAATCCANTTCARAAATTA 4127
281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
4128 GCAACACAGCGAGGCTAACGTAACTAAGTCAACAGGTACAAATTTGGTGTACGATTTC 4187
301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLeuGln 320
4188 GGTAATACCGTAAAT-----GTTACGGCAACCGCTGGC 4220
321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
4221 GATTAAACAGTTGGGAATGGCGCAGAA-----ATTAATGCGACAGAA 4262
341 SerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
4263 GGAGCTGCAACCTTAACCTACATCATCGGCGCAATTA-----ACT 4301
361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTySerGlnAsp 380
4302 ACCGAGTAGTGTACACATTAATCTACGCCAAGGTCAGGTAAATCTTTCAGTCAGGAT 4361
381 GlyTyProGlnGlyAspLeu-----ValAspValThrIleThrSerGlyLysLeu 398
4362 GGTAGCGTTGCGGAAGTATTAAATGCGCCCAATGTGACACTAAATACTACAGGCACCTTA 4421

399 -----GlnGlyLys-----TyrSerAsnSerGln 406
4422 ACTACCGTGAAGGTTCAACATTAATGCAACGCGGTACCTTGGTTATTAAACGCAAA 4481
407 ValValAspPheTyThrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
4482 GACGCTGAGCTAAATGCGCAGCATTTGGTAAACACACAGTGGTAAATGCAACCAACGCA 4541
427 GluGlyAsnAsnHisTyThrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
4542 AATGCTCCGCGCAGGTAATCGCAGCACTCAAGC-----AGAGTGACATCACT 4592
447 GlyThrSerAsnTyThrLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
4593 GGG-----GATTTAATCACATAAATGATTA-----AATATC---ATT 4628
467 SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
4629 TCMAAAACGCTATAAACACCGTACTGTGA---AAAGCGTTAAATGTGATGTGAATAC 4685
487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
4686 ATTCAACCGGTTATAGCAACGCTAGATGAAGTAATGGAAGCAACGCG 4733

RESULT 22
US-08-617-697-1
; Sequence 1, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-617-697-1

Alignment Scores:

Pred. No.: 7,398-08 Length: 5116
 Score: 173.50 Matches: 110
 Percent Similarity: 37.30% Conservative: 75
 Best Local Similarity: 22.18% Mismatches: 210
 Query Match: 6.71% Indels: 101
 DB: 2 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-617-697-1 (1-5116)

```

QY 30 AsnThrIleGlyTyrIleGlyGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3480 AATATTTCAGGTTTCAATAAAGCAGAGATTACAGCTAAAGAT-----GGTAGTGATTTA 3533

QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3534 ACTATTGGTAACACCAATAGTCTGATGATAAT----- 3569

QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3570 ---GCCAAAAAAGTAACCTTT-----AACCCAGTTAAAGAT 3602

QY 89 LeuAlaIleGlyGlyGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3603 TCAAAAATCTCTGCTGACGGT---CACAAAGTGACACTACACAGCAAGTGAA----- 3653

QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3654 ACATCCGGTAGTAATAACACACTGAAGATGACGTGACATAT---GCCGGCTTAAC 3710

QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3711 ATCGATGCAAAAATGTAAACAGTAACCAATATTAATCTCTCAAAAGCAGTGAGCATC 3770

QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3771 -----TCTGCACAAAGTGGAGAAATTAACCACTAAACAGGTACACCACT 3815

QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3816 AACGCAACCACTGGTAACCTGAGATTAACCGCTCAACACAGTAGT-----ATC 3863

QY 189 PheAlaLeuLeuGluSerThrLysGlyAsnGlyThr-----ProProIle 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3864 CTAGGTGAATGTAGTCCAGCTCTGGCTCTGTAACTTACTGCAACCGAGCGCTCTT 3923

QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3924 GCTGTAAGCAATATTTCG-----GGCAACACC 3950

QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3951 GTTACTGTGTACTGCAAAAT---AGCGGTGCATTAAACCACTTTGGCAGGCTCTACAAATAAA 4007

QY 244 TyrLeuValAlaAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4008 GGAACCGAGGTGTAACCACTTCAAGTCATCAGCGGATTCGGGTAGCATTTCTGGT 4067

QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4068 GGCACATAGAGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAATTCAAAATATAA 4127

QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4128 GCAACACAGCGGAGGTACGTAAACAGTGCAACAGGTACCAATTTGGTGTAGCATTTCC 4187

QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4188 GGTAATAACGGTAAAT-----GTTACCGCAAAACGGTGGC 4220

QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4221 GATTTAACGTTGGAAATGCCGAGAA-----ATTAATGCCGACAGAA 4262
  
```

```

QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4263 GGAGCTGCACACCTTAACATACATCATCGGCAAAATTA-----ACT 4301

QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4302 ACCGAAGCTAGTTCACACATTACTTCCGCCAAGGTCAGGTAATACTTTTCAGTCAGGAT 4361

QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4362 GGTAGCGTTGCAGGAAGTATTAAATGCCGCCAATGTGACACTAAATACTACAGGCACTTTA 4421

QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4422 ACTACCGTGAAGGGTTCAAAACATTAAATGCAACCGCGGTACCTTGGTTATTAAACGCAAAA 4481

QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4482 GAGCGTACGCTAAATGGCGCAGCATTTGGTAACACACAGTGGTAATACCAACCAACGCA 4541

QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4542 AATGGTCCGCGAGCGGTAATTCGCAACACCTCAAGC-----AGAGTGAACATCACT 4592

QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAlaGlnLeuGluThrSerAsnValAspMet 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4593 GGG-----GATTTAATCAATAATGGAATTA-----AATATC---ATT 4628

QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4629 TCAAAAACCGGTATAAACACCGCTACTGTTA---AAAGCGGTAAATATGATGTGAATAC 4685

QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4686 ATTCAACCGGTATAGCAAGCGGTAGATGAAGTAATTTGAAGCGAAACGC 4733
  
```

RESULT 23

US-08-719-641-1
 ; Sequence 1, Application US/08719641
 ; Patent No. 6218141
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 16-SEP-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-625
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5116 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-719-641-1

Alignment Scores:

Pred. No.: 7,398-08 Length: 5116
 Score: 173.50 Matches: 110
 Percent Similarity: 37.30% Conservative: 75
 Best Local Similarity: 22.18% Mismatches: 210
 Query Match: 6.71% Indels: 101
 DB: 3 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-719-641-1 (1-5116)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
 DB 3480 AATATTTTCAGGTTTCAATAAAGACGAGATTACAGCTAAAGAT-----GGTAGTGATTAA 3533
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
 DB 3534 ACTATTGGTAACACCAAGTAGTGCTGATGCTACTAAT----- 3569
 QY 59 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
 DB 3570 ---GCCAAAAGTAACCTTT----- 3602
 QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108
 DB 3603 TCAAAATCTCTGCTGACGGT---CACAGGTGACACTACACGAAAGTGGA----- 3653
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 DB 3654 ACATCCGGTAGTATAACACACTGAAGATAGCAGTACCAATAT---GCCGGCTTAAC 3710
 QY 129 LeuMetGlySerArgIleSerAsnProAsnIleLysLysGluThrLeuGluProIle 148
 DB 3711 ATCGATGCACAAAATGTACAGTAAACACANTATTACTTCTCACAAAGCAGTGAGCATC 3770
 QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
 DB 3771 -----TCTGCACAAAGTGGAGAAATTAACCACTAAACACAGGTACAAACCAATT 3815
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 DB 3816 AACGCAACCACTGGTAACGTGAGATACCCCTCAACAGTAGT-----ATC 3863
 QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle 203
 DB 3864 CTAGGTGGAATAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGGGCGCTCT 3923
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
 DB 3924 GCTTAACAAATATTTCG-----GCCACACC 3950
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
 DB 3951 GTTACTGTGTCAAAAT---AGCGGTGCATTACCACTTTGGCAGGCTCTACAAATAAA 4007
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
 DB 4008 GGAACCGAGAGTGTAACCACTTCAAGTCATCAGCGGATATCGCGGTAGTATTTCTGGT 4067

QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 DB 4068 GGCACAGTAGAGGTTAAAGCAACCGAAAGTTTAAACCACTCAATCCAATTCAAAATTTAA 4127
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
 DB 4128 GCAACAACACGCGGAGGTAAAGTCAAGTCAACAGCAGTCAATTTGGTGTGACGATTTC 4187
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 DB 4188 GGTAAATACGGTAAT-----GTACGGCAACCGTGGC 4220
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
 DB 4221 GATTTAACAGTTGGGAATGGCGCAGAA-----ATTATGCGACAGAA 4262
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 DB 4263 GGAGCTGCACCTTAACCTATCATCGGCAAAATTA-----ACT 4301
 QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
 DB 4302 ACCGAAGCTAGTTTACACCAITACTTTCAGCCAGGTCAGGTAATCTTTCAGCTCAGAT 4361
 QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
 DB 4362 GGTAGCGTTGCAAGAGTATTAAATGCGCCCAATGTGACACTAAATACACAGCACTTTA 4421
 QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
 DB 4422 ACTACCGTGAAGGTTCAAAACATTAAATGCAACGCGGTACCTTGGTTATTAAACGAAA 4481
 QY 407 ValValAspPheThrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
 DB 4482 GACGCTGAGCTAAATGCGCAGCAGCATTTGGTAAACACACAGTGTAAATCAACCAACGCA 4541
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
 DB 4542 AATGGCTCGGACGCGTAATCGCAGCACTCAAGC-----AGAGTGAACATCACT 4592
 QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
 DB 4593 GGG-----GATTTAATCAATAAATGGATTA-----AATATC---ATT 4628
 QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
 DB 4629 TCAAAAACGCTATAAACACCGTACTGTTA---AAAGGCGTTAAATTTGATGTGAATATC 4685
 QY 487 ValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
 DB 4686 ATTCACCGGCTATAGCAAGCGTAGATGAAGTAATTTGAAGCGGAACGC 4733

RESULT 24

US-09-206-942-66
 Sequence 66, Application US/09206942
 Patent No. 6432869
 GENERAL INFORMATION:
 APPLICANT: Loosmore, Sheena M.
 APPLICANT: Yang, Yan-ping
 APPLICANT: Klein, Michel H.
 TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 TITLE OF INVENTION: Molecular Weight Proteins
 FILE REFERENCE: 1038-861 M1S:jb
 CURRENT APPLICATION NUMBER: US/09/206,942
 CURRENT FILING DATE: 1998-12-08
 EARLIER APPLICATION NUMBER: 09/167,568
 EARLIER FILING DATE: 1998-10-07
 NUMBER OF SEQ ID NOS: 95
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 66
 LENGTH: 5116
 TYPE: DNA
 ORGANISM: Haemophilus influenzae

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9171 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-038-682-5

Alignment Scores:

Pred. No.: 1.92e-07 Length: 9171
 Score: 173.50 Matches: 110
 Percent Similarity: 37.30% Conservativity: 75
 Best Local Similarity: 22.18% Mismatches: 210
 Query Match: 6.71% Indels: 101
 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-038-682-5 (1-9171)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
 DB 3430 AATATTTTCAGGTTTCATTAAGCAGAGATTACAGCTAAAGAT-----GGTAGTGATT 3483
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
 DB 3484 ACTATTGGTAACCAATAGTCTGTGAGTACTAAT----- 3519
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
 DB 3520 ---GCCAAAAGTAACCTTT-----AACAGGTTAAAGAT 3552
 QY 89 LeuAlaIleGlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThr 108
 DB 3553 TCAAAATCTCTGCTGAGGCT---CACAAAGTGACACTACACGCAAAAGTGAA----- 3603
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 DB 3604 ACATCCGCTAGTATTAACACACTCAAGATAGACAGTACACATAAT---GCCGCTTAAC 3660
 QY 129 LeuMetGlySerArgIleSerAsnProAsnIleLysLysGluThrLeuGluProIle 148
 DB 3661 ATCGATGCAAAATGTAACAGTAGTAACACAAATATTACTTCTCACAAAGCAGTGAGCATC 3720
 QY 149 GlnLeuAspPheAsnProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
 DB 3721 -----TCGCGACAGTGGAGAAATTACCACTAAACAGGTACACCAT 3765
 QY 169 AsnAlaValAlaLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
 DB 3766 AACGCAACCACTGTAACGTGAGATAACCGCTCAAAACAGGTAGT-----ATC 3813
 QY 189 PheAlaLeuLeuGluSerTyrLysGlyAsnGlyThr-----ProProIle 203
 DB 3814 CTAGGTGAAATTGATCCAGCTCTGGCTCTGTAACTTACTCAACCGGCGGCTCTT 3873
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
 DB 3874 GCTGTAGCAATATTTCG-----GGCAACACC 3900
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
 DB 3901 GTTACTGTATTGCAAT---AGCGGTGCATTAACCACTTGGCAGGCTCTACATATA 3957
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
 DB 3958 GGAACCGAGAGTGAACCACTTCAAGTCAATCAGCGCATATCGCGGTACGATTCTGTT 4017
 QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 DB 4018 GCACAGTAGAGTTAAGAACCAAGATTACCACTCAATCCCAATCAAAATATA 4077
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaThrProGlnPro 300

DB 4078 GCAACAACAGCGGAGGCTAACGTAACAGTGCAACAGGTACAAATTGGTGGTAGATTCC 4137
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 DB 4138 GGTAATACGGTAAAT-----GTTACGCAACACGCTGGC 4170
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrrPalalGlyAlaProAla 340
 DB 4171 GATTTAACAGTTGGGAATGGCGCAGAA-----ATTATGCGCAGAA 4212
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
 DB 4213 GGAGCTGCAACCTTAACATACATCGGCAATAA-----ACT 4251
 QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
 DB 4252 ACCGAAGCTAGTTCACACATTCTTCACCAAGGTCAGGTAAATCTTTCAGCTCAGGAT 4311
 QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
 DB 4312 GGTAGCGTTGCAGGAAGTATTAAATGCCGCAATGTGACACTAAATACTACAGCACTTTA 4371
 QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
 DB 4372 ACTACCGTAGGAGGTTCAACATTAATGCAACGCGGTACCTTGGTTATTATACGCAAAA 4431
 QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
 DB 4432 GACGCTGAGCTAAATGCGCAGCATTTGGTAAACACACAGTGTGTAATCAACCAACGCA 4491
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
 DB 4492 AATGCTCCCGCAGCGTAATCGGCAACCTCAAGC-----AGAGTGAACATCACT 4542
 QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
 DB 4543 GGG-----GATTTAATCAACAATAATGGATTA-----AATATC---ATT 4578
 QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
 DB 4579 TCAAAACCGGTATAAACACCGTACTGTTA---AAAGCGCTTAAATTTGATGGAATAC 4635
 QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
 DB 4636 ATTCAACCGGTTATAGCAACGCTAGATGAAGTAATTTGAACGGAACGC 4683

RESULT 26

US-08-302-832-5
 ; Sequence 5, Application US/08302832
 ; Patent No. 5603938

GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/302,832
 ; FILING DATE: 16-SEP-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:

```

/ APPLICATION NUMBER: GB 9205704.1
/ FILING DATE: 16-MAR-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US pct/us93/02166
/ FILING DATE: 16-MAR-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Berkstresser, Jerry W
/ REGISTRATION NUMBER: 22,651
/ REFERENCE/DOCKET NUMBER: 1038-404
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 415-0810
/ TELEFAX: (703) 415-0813
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 9171 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-302-832-5

Alignment Scores:
Pred. No.: 1,92e-07 Length: 9171
Score: 173.50 Matches: 110
Percent Similarity: 37.30% Conservativity: 75
Best Local Similarity: 22.18% Mismatches: 210
Query Match: 6.71% Indels: 101
DB: 1 Gaps: 24

US-10-009-823a-1 (1-502) x US-08-302-832-5 (1-9171)
QY 30 AsnThrLeuGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
DB 3430 AATATTTCAGGTTTCAATAAGCAGACGATTACAGCTAAAGAT-----GGTAGTGATTTA 3483
QY 50 AlaLeuGlySerThrGlySer-----GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
DB 3484 ACTATTGGTAACCAATAGTGTGATGTTGTTACTAAAT----- 3519
QY 69 GlySerValArgThrIlePheThrGlnGlnValAlaPheGluProGlyAsnSerValThrAsp 88
DB 3520 ---GCCAAAAGTAACCTTT-----AACCAAGTTAAAGAT 3552
QY 89 LeuAlaIleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisTyrThr 108
DB 3553 TCAAAAATCTCTGCTGACGGT---CACAGGTGCACACTACACAGCAAGTGGAA----- 3603
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
DB 3604 ACATCCGGTAGTAATAACACACTGAAGTAGCAGTGACAAATAT---CCCGCTTAATC 3660
QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysGlyLeuThrLeuGluProIle 148
DB 3661 ATCGATGCAAAAATGTAACAGTAACAAACAATATATTCTTCACAAAGCAGTGAGCATC 3720
QY 149 GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu 168
DB 3721 -----TCTGGCAGTGGAGAAATTAACCACTAAACAGGTCAACCACTATT 3765
QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr 188
DB 3766 AACGCAACCACTGCTAAACGTTGAGATAACCGCTCAACACAGGTAGT-----ATC 3813
QY 189 PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProIle 203
DB 3814 CTAGGTGGAAATGAGTCCAGCTCTGCTCTGTAACACTTACTGCAACCGGCGCTCTT 3873
QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
DB 3874 GCTGTAAAGCAATATTTCG-----GGCAACACC 3900
QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerThrPheGlu 243
DB 243 ----- 243

```

```

DB 3901 GTTACTGTCTACTGCAAAAT---AGCGGTGCATTAACCACTTTTGCAGGCTCTCAATAATAA 3957
QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
DB 3958 GGAACCGAGAGTGAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTTCTGGT 4017
QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyLeuLeuLys 280
DB 4018 GGCACAGTAGAGTTAAAGCAACCAAGTTTAAACCACTCAATCCATTTCAAAATATAA 4077
QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
DB 4078 GCAACACAGCGGAGCTTAACCAAGTGAACAGGTACAAATGGTGGTACGATTTCC 4137
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
DB 4138 GGTAAATACGGTAAAT-----GTTACGCAACCGCTGGC 4170
QY 321 ProLeuThrLeuAspPheGlyLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
DB 4171 GATTAAACAGTTGGGAATGGCGCAGAA-----ATTAATCGCAGCAA 4212
QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360
DB 4213 GCGAGCTGCACCTTAACCTACATCATCGCGCAATTA-----ACT 4251
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp 380
DB 4252 ACCGAAGCTAGTTTACACATTTACTTACGCCAAGGCTCAGGTAAATCTTTTCAGCTCAGAT 4311
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
DB 4312 GGTAGCGTTGAGGAAGTATTAAATGCCGCCAATGTGACACTAAATACTACAGGCACCTTTA 4371
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
DB 4372 ACTACCGTGAAGGTTTCAAACTTAATGCAACCAAGCGGTACCTTGGTTATTAAACGCAAA 4431
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
DB 4432 GACGCTGAGCTAAATGCGCGAGCAATTTGGTAAACACACAGTGGTAAATGCAACCAACGCA 4491
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
DB 4492 AATGGCTCCGCGCAGCGTAATCGCGACAAACCTCAAGC-----AGAGTGAACATCACT 4542
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
DB 4543 GGG-----GATTTAATCACAATAAATGGATTA-----AATATC---ATT 4578
QY 467 SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
DB 4579 TCAAAAACCGGTATAACACCGTACTGTTA---AAAGCGGTAAATGTATGTGAATATAC 4635
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
DB 4636 ATTCAACCGGTTAGCAAGCGTAGATGAAGTAATTGAAGCGAAACGCC 4683

RESULT 27
US-08-530-198-5
/ Sequence 5, Application US/08530198
/ Patent No. 5869065
/ GENERAL INFORMATION:
/ APPLICANT: BARENKAMP, STEPHEN J
/ APPLICANT: ST. GENE III, JOSEPH W
/ TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
/ NUMBER OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
/ NUMBER OF SEQUENCES: 8
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Shoemaker and Mattare, Ltd
/ STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
/ CITY: Arlington

```


;
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/469,880
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 16-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9171 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-469-880-5

Alignment Scores:
 Pred. No.: 1.92e-07 Length: 9171
 Score: 173.50 Matches: 110
 Percent Similarity: 37.30% Conservative: 75
 Best Local Similarity: 22.18% Mismatches: 210
 Query Match: 6.71% Indels: 101
 DB: 2 Gaps: 24

US-10-009-823A-1 (1-502) x US-08-469-880-5 (1-9171)

Qy	129	LeuMetGlySerArgIleSerAsnProAsnIleGlyLeuThrLeuGluProIle	148
Db	3661	ATCGATGCAAAAAATGTAACAGTAACAAATATTACTTCTCACAACAGTACATC	3720
Qy	149	GlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeu	168
Db	3721	-----TCTCGGACAGTGGAGAAATTACCCTAAACAGGTCAACCAT	3765
Qy	169	AsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyr	188
Db	3766	AACGCAACCACTGGTAACCGTGAGATAACCGCTCAACAGGTAGT-----ATC	3813
Qy	189	PheAlaLeuLeuGluSerTrpLysGlyAsnGlyThr-----ProProIle	203
Db	3814	CTAGGTGGAATTGAGTCCAGCTCTGGCTCTGTAAACACTTACTGCAACCGAGCGCTCTT	3873
Qy	204	SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer	223
Db	3874	GCTGTAAAGCAATATTTCG-----GGCAACACC	3900
Qy	224	HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu	243
Db	3901	GTTACTGTACTGCAAAAT---AGCGGTGCATTAAACCACTTTGGCAGGCTCTACAAATAA	3957
Qy	244	TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla	263
Db	3958	GGAACCGAGAGTGTAAACCACTTCAAGTCAATCAGCGGATATCGCGGTACCATTTCTGGT	4017
Qy	264	GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyLeuLeuLys	280
Db	4018	GGCACAGTAGAGTTAAAGCAACCCGAAAGTTTAACTCAATCCAAATTCAAAATTTAA	4077
Qy	281	AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGlnPro	300
Db	4078	GCACAACAGCGGAGGTAAGTAAAGTGAACAGGTGCAACAGGTACATTTGGTGGTACATTC	4137
Qy	301	AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLeuGln	320
Db	4138	GGTAATACGGTAAAT-----GTTACGGCAACCGTGGC	4170
Qy	321	ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTyrAlaGlyAlaProAla	340
Db	4171	GATTTACAGTTGGATGGCGCAGAA-----ATTAATGCCACAGAA	4212
Qy	341	SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr	360
Db	4213	GGAGCTGCAACCTTAACCTACATCATCGGCAAAATTA-----ACT	4251
Qy	361	SerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAsp	380
Db	4252	ACCGAAGCTAGTTACACATTTACTTCCAGCCCAAGGGTCAGGTAAATCTTTCAGCTCAGAT	4311
Qy	381	GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGlyGlyLysLeu	398
Db	4312	GGTAGCGTGGCAGGAGTATTATGCGCCCAATGTGACACTAATACTACAGGCACTTTA	4371
Qy	399	-----GlnGlyLys-----TyrSerAsnSerGln	406
Db	4372	ACTACCTGAGGGTTCAACATTTAATGCAACCGCGTACCTTGGTTATTAAACGCAAA	4431
Qy	407	ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg	426
Db	4432	GACCTGAGCTAAATGGCGGAGGATTTGGTTAACCACAGTGGTAAATGCAACCAACGCA	4491
Qy	427	GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro	446
Db	4492	AATGGCTCCGCGACGCTAATCGCGCAACCACTCAAGC-----AGAGTGAACATCACT	4542
Qy	447	GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlnThrSerAsnValAspMet	466
Db	4543	GGG-----GATTTAATCACAATAATGGATTA-----AATATC---ATT	4578
Qy	467	SerArgGluMetValAsnMetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSer	486

[illegible]

```

RESULT 29
US-08-728-470-5
; Sequence 5, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattcare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-728-470-5

Alignment Scores:
Pred. No.:      Length:      9171
Score:          Matches:      110
Percent Similarity: 37.30%    Conservative: 75
Best Local Similarity: 22.18%   Mismatches: 210
Query Match:      Indels:     101
DB:               Gaps:       24

US-10-009-823A-1 (1-502) x US-08-728-470-5 (1-9171)

QY      30 AsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49
      ||| ||||| : : : : :
Db      3430 AATATTTTCAGGTTCATAAAGCAGACATTACAGTAAGAAT-----GGTAGTGATT 3483
QY      50 AlaIleGlySerThrGlytSer---GlnGlyProAsnGlnAlaGlyMetGlyValGlnVal 68

```

Db 4372 ACTACCGTGAAGGGTTCAAAACATTAAATGCAACACCGGTACCTGGTTATTAAACGCAAAA 4431
 QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
 Db 4432 GACGCTGACATAATGGCCGACATGGGTGACACACACAGCTGGTAAATGCAACACGCA 4491
 QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
 Db 4492 AATGGCTCCGGCAGCGTAAATCCGCAACACCTCAAGC-----AGAGTGAACATCACT 4542
 QY 447 GlyThrSerAsnTyrGlyIysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
 Db 4543 GGG-----GATTAAATCAATAATGGATT-----AATATC---ATT 4578
 QY 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerIysSer 486
 Db 4579 TCAAAAACCGGTATAAACAACCGTACTGTGA---AAAGCGGTAAATAATGTGATGTAATAC 4635
 QY 487 ValThrThrAlaAspThrMetLeuGlnIysAlaLeuGluLeuIysArg 502
 Db 4636 ATTCAACCGGTATAGCAAGCGGTAGATGAAGTAATGAAGCGCAACGCG 4683

RESULT 30

US-08-617-697-5
 ; Sequence 5, Application US/08617697
 ; Patent No. 5977336
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/617,697
 ; FILING DATE: 01-APR-1996
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 05-OCT-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-557
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9171 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-617-697-5

Alignment Scores:

Pred. No.: 1.92e-07 Length: 9171
 Score: 173.50 Matches: 110

Percent Similarity: 37.30% Conservative: 75
 Best Local Similarity: 22.18% Mismatches: 210
 Query Match: 6.71% Indels: 101
 DB: 2 Gaps: 24
 US-10-009-823a-1 (1-502) x US-08-617-697-5 (1-9171)
 QY 30 AsnThrIleGlyTyrIysGlnGlnValPheGlnAspLeuPheSerGlnAspLeu 49
 Db 3430 AATATTTCAGTTTCATTAACAGCAGATTACAGTAAAGAT-----GGTAGTGATTTA 3483
 QY 50 AlaIleGlySerThrGlySer---GlnGlyProAsnGlnAlaGlyMetGlyAlaGlnVal 68
 Db 3484 ACTATTGGTAACCAATATAGTCTGATGTACTAAT----- 3519
 QY 69 GlySerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAsp 88
 Db 3520 ---GCCAAAAAGTAACCTTT-----AACAGGTTAAAGAT 3552
 QY 89 LeuAlaIleGlyGlyIysGlyPheGlnValThrLeuGluAspIysValHisTyrThr 108
 Db 3553 TCAAAAATCTCTGCTGACCGT---CACAAAGGTGACACTACACAGCAAAAGTGAA 3603
 QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
 Db 3604 ACATCCGGTAGTAATAACAACTGAAGTAGCAGTGAATAAT---GCCGGCTTAAC 3660
 QY 129 LeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIle 148
 Db 3661 ATCGATGCAAAAAATGTAACAGTAACAAACAATATTACTTCTCAAAAGCAGTGGAGCATC 3720
 QY 149 GlnLeuAspPheAsnAspProThrValAlaIysSerProAlaIysThrSerThrAlaLeu 168
 Db 3721 -----TCTGCCGACAGTGGAGAAATATACCACATAAACAGGTACAAACCAAT 3765
 QY 169 AsnAlaValValAsnLeuGlyAspSerThrAspIysThrGlnSerGluAlaAsnProTyr 188
 Db 3766 AACGCAACCACTGGTAACGTGGAGATACCTCCCTCAACAGGAGTAGT-----ATC 3813
 QY 189 PheAlaLeuLeuGluSerTrpIysGlyAsnGlyThr-----ProProIle 203
 Db 3814 CTAGGTGGAATTTAGTCCAGCTCTGGCTCTCTAAACACTTACTGCAACCGAGGGCGCTCT 3873
 QY 204 SerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSer 223
 Db 3874 GCTGTAAACAATATTCG-----GGCAACACC 3900
 QY 224 HisAspIleThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGlu 243
 Db 3901 GTTACTGTACTGCAAAAT---AGCGGTGCATTAACCACTTTGGCAGGCTCTACAATATAA 3957
 QY 244 TyrLeuValAlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAla 263
 Db 3958 GGAACCGGAGGTGAACCACTTCAAGTCAATCAGCGCATATCGCGGTACCATTTCTGGT 4017
 QY 264 GlyLeuLeu-----MetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLys 280
 Db 4018 GGCAGTAGAGGTAAAGCAACCGAAAGTTTAACCACTCAATCCAATTCAAAATATAA 4077
 QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrIysAspLeuAsnAlaTrpGlnPro 300
 Db 4078 GCAACCAACAGCGGAGGTAAAGTAAACAGGTGCAACAGGTACAAATTTGGTGTACGATTTC 4137
 QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln 320
 Db 4138 GGTAAATCGGTAAAT-----GTTACGGCAACCGCTGGC 4170
 QY 321 ProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
 Db 4171 GATTTAAACAGTTGGGAATGGCGCAGAA-----ATTAATGCGACAGAA 4212
 QY 341 SerAlaAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThr 360


```

QY 281 AsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnPro 300
Db 4078 GCAACAACAGCAGCGAGTAAACGTAACAAGTGCACACAGTAAATGGTGGTACGATTTC 4137
QY 301 AlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLeuGln 320
Db 4138 GGTATACGTTAAAT-----GTTACGCAACGCTGGC 4170
QY 321 ProLeuThrLeuAspPheGlyLeuLysSerGlnGlnAsnMetTrpAlaGlyAlaProAla 340
Db 4171 GATTTAAACAGTGGCAATGGCGCAGAA-----ATTAATGCCACAGAA 4212
QY 341 SerAlaAlaAlaGlyThrAspLeuGlyLysLeuProSerMetMetProIleGlnThr 360
Db 4213 GGAGCTGCAACCTTAACATACATCATCGGCGCAATTA-----ACT 4251
QY 361 SerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp 380
Db 4252 ACCGAAGCTAGTTACACATTAATTCAGCCAGGTCAGGTAATCTTCAGCTCAGGAT 4311
QY 381 GlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSerGluGlyLysLeu 398
Db 4312 GGTAGCGTTGCGAGGAATTAATGCGCCCAATGTGACACTAAATACTACTACAGGCACCTTA 4371
QY 399 -----GlnGlyLys-----TyrSerAsnSerGln 406
Db 4372 ACTACCGTGAAGGTTCAACATTAATCAACACGCGTACCTGGTTATTATTAACGCAAAA 4431
QY 407 ValValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgArg 426
Db 4432 GACGCTGAGCTAAATGGCGCAGCATTTGGTAAACACACAGCTGGTAAATCAACCAACGCA 4491
QY 427 GluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPro 446
Db 4492 AATGGCTCCGCGAGGTAATCGACACACCTCAAG-----AGAGTGAACATCACT 4542
QY 447 GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet 466
Db 4543 GGG-----GATTTAATCACAATAATGGATTA-----AATATC---ATT 4578
QY 467 SerArgGluMetValAsnMetIleIleIleGlnhArgGlyPheGlnMetAsnSerLysSer 486
Db 4579 TCRAAACAAGTATAAACACCGTACTGTTA---AAAGCGGTAAATATGATGTAATAC 4635
QY 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db 4636 ATTCAACCGGTATAGCAAGCGTAGATGAAGTAATTGAAGCAACGC 4683

```

RESULT 22

```

US-09-453-702B-39
; Sequence 39, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:

```

```

; APPLICANT: Blattner, Frederick R.
;            Perna, Nicole T.
;            Plunkett, Guy
;            Welch, Rod
;            Flunkett, Guy

```

```

; TITLE OF INVENTION: No. 6365723e1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Quarles & Brady
; STREET: 1 South Finckney Street
; CITY: Madison
; STATE: WI

```

```

; COUNTRY: US
; ZIP: 53701-2113

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-Dec-1998
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 39:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 25165
; STRANDEDNESS: double

```

```

; TYPE: nucleic acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:

```

```

US-09-453-702B-39
Alignment Scores:

```

```

Pred. No.: 8-27e-06 Length: 25165
Score: 165.00 Matches: 145
Percent Similarity: 33.70% Conservative: 71
Best Local Similarity: 22.62% Mismatches: 189
Query Match: 6.38% Indels: 236
DB: 3 Gaps: 35

```

```

US-10-009-823A-1 (1-502) x US-09-453-702B-39 (1-25165)

```

```

QY 8 GlyAlaThrGlyMetLysThrHis-----SerThrGlyLeuGlyThrValSer 23
Db 13018 GGAACACGCGGCGGCTACCCATCAGGTGACGTCATACCGGCTCCGACCATACC 13077
QY 24 -----AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGln 37
Db 13078 TTTAACGCCATCAGCGCGGATAACATCTGAACCGCGATGAAGGCG----- 13125
QY 38 GlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIle-----GlySerThrGly 55
Db 13126 -----CAGCGCTTGACCATCAGCGCGGCGGAGTACGCGG 13158
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 13159 -----CTGCGCAGCGGCGGCGGAGTCT-----ACGTCACG 13188
QY 76 ThrGlnGly-----AlaPheGluProGlyAsn----- 84
Db 13189 CTCACGCGTCAACACTACAGCGCCACACCGACGATCGGCGCAACTGACCTTAACCGTG 13248
QY 85 SerValThrAspLeuAlaIleGlyLysGlyPheGlnValThrLeuGluAspLys 104
Db 13249 CCGGTGACGCGATCGCGGCGATTAGTCAGCCCACTATACGTCAGCGCCGCGCCACC 13308
QY 105 ValHisTyrThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspPro 124
Db 13309 -----AGTGCAGCAGGCAACACCGCCAGCAGCGGAGGAAATTTACTGGTCGAC--- 13356
QY 125 SerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysGluThr 144
Db 13357 -----AGCGCGCTGCGGCGGAGTCAACATCAACACACC 13386
QY 145 LeuGluProIleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThr 164
Db 13387 GTGGCA-----GGCGACGATATTATCAACCGCGCGGAGCGGGGCC 13428
QY 165 SerThrAlaLeuAsnAlaValValAsn-----LeuGlyAspSerThrAspLysThr 181
Db 13429 GATCAACCATCAGCGGGGTGGTGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13488

```


QY 182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTyrLysGlyAsnGlyThrPro 201
Db 13489 CTGGGC---GGGAACACTTACACCGCTACGTACAG-----13521
QY 202 ProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGln--- 220
Db 13522 -----AGCACTTAAGCTGGAGCGTCCGACGCGGATCTCCAGCG 13569
QY 221 ---GlyAsnSerHis---AspIleThrValTyrPheAspGlyAlaProSerThrGly 238
Db 13570 TTGGGCAATGGGATTGACCAATTACCGCTCGGTCCAAACAGCTAATGGCAACACCGG 13629
QY 239 SerLysThrPheGluTyrLeuVal-----246
Db 13630 ACGGCACGCGGATATACCATTCATGCGCAACCTGCGGGGCTGCGGTAGATACCGTG 13689
QY 247 -----AlaMetAsnProSerGluAspGlySerAlaAlaSerGlyThr---Asp 261
Db 13690 GCGGCGGATGATATCGTCAACAGCATCGACGCGGAGCGGCTGGTGTATCACCGCGGC 13749
QY 262 SerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeuLysAsn 281
Db 13750 AGTAGCGGCTGAATGCAGGTCTGTCTGACGGTTACCATCAACAGT-----13797
QY 282 MetThrAlaPheThrProThr-----GlySerAlaThr 292
Db 13798 ---GTGGCGTATTCGCCACCGTGCAGCGGACGGAAGCTGGAGCTGGCATTCGCGCG 13854
QY 293 LysAspLeuAsnAlaTyrGlnProAlaProLeu-----ValAsnGlyLeuProGln 309
Db 13855 GCAACGCTCAGCGCTGCGCTGCGGGCGGTTAACCGTGGAGGTAGACGG-----CAA 13908
QY 310 PheSerAlaAsnPheValGlyAlaGlyLeuGlnProLeuThrLeuAspPhe----- 326
Db 13909 AGCAGCGCAATAAACCCAGTCAGCGTACGCATCCGTCACCGTTCGATTTAACGGCGGTG 13968
QY 326 -----326
Db 13969 GCAATCAGCATCAACACCGTTGCCAGCGACGACGTGATTAAACGCGCGAGAAAGGCACC 14028
QY 327 -----GlyIleLysSerGlnGlnAsnMet-----334
Db 14029 AATCTGACTCTTCGGCGAGTACACGCGGATTCAGAGCGGGCAAAACCGTCACCGTCACT 14088
QY 335 TrpAlaGlyAla-----ProAlaSerAlaAlaIleGlyThr-----347
Db 14089 TTTGCGGTAAACCTACACTGCAAGCGTCCGCGCAACGCGAGCTCGAGTGTAAACGTT 14148
QY 348 -----AspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsn 364
Db 14149 CCGCGCGGAGATCTGGCAACTCTGCCAGAGCGCGCGCAATGTGACGAGCGCGGTAGC 14208
QY 365 SerThrAlaArgAsnGlySerSerThrArgArgTyrSerGlnAsp-----380
Db 14209 ACGCGAGCGGTAAACAGTGCCTCGCGACCCATCGGTATAGCTTGACGCGCGCGCGCG 14268
QY 381 -----GlyTyr 392
Db 14269 ACGCTCACCATTACACCATCGCCAGCAGCATATCTTAAACGCGCGAGAGCGCGAAGC 14328
QY 383 Pro-----GlnGlyAspLeuValAspValThrIle 392
Db 14329 CGCTCACCATCAGCGGACGACGACCGCGCAACCGGGGACGCGTACCGTACCCCTT 14388
QY 393 -----ThrSerGluGlyLysLeuGln-----GlyLysTyrSerAsnSerGlnVal 407
Db 14389 AACGGCGCAACCTACACCGGCACTGTGCAGCGGACGCTAGCTGGAGCGTCAAG-----14442
QY 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427
Db 14443 -----GTTCCC-----ACTTCAGCCCTCGGCGCGCTCAACGCA 14475

QY 428 GlyAsnAsnHistyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGly 447
Db 14476 AGCAATTACCGTCAGCGCACCGCTCAATGAC-----AAAGCGGCAACCCCCGCG 14526
QY 448 ThrSerAsnTyrGly-----LysLeuSerValAsnGlnLeu 459
Db 14527 AGCGCAGCCATAATCTGGCGGTAGACACACCGCGCGGTCTCTCAACATTAAACCGGTG 14586
QY 460 GluThrSerAsnValAspMetSerArgLysMetValAsnMetIleIleGlnArg--- 478
Db 14587 GCGGCGGATGACATCATCAACGATGCGCAACATGCGCAGCGCTGGTGTCTCCGGCACC 14646
QY 479 -----GlyPheGlnMetAsnSerLysSerValThr 488
Db 14647 AGTAGCGCGGGAAGCGGCGATGTGTGTGAGCGTGTGTCTCAACGCAAAACCTACACC 14706
QY 489 Thr 489
Db 14707 ACC 14709
RESULT 33
US-09-252-991A-13774/c
; Sequence 13774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13774
; LENGTH: 4188
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13774

Alignment Scores:

Pred. No.:	1,52e-06	Length:	4188
Score:	160.00	Matches:	111
Percent Similarity:	35.67%	Conservative:	62
Best Local Similarity:	22.89%	Mismatches:	180
Query Match:	6.13%	Indels:	132
DB:	4	Gaps:	27

US-10-009-823A-1 (1-502) x US-09-252-991A-13774 (1-4188)

QY 53 SerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerVal--- 71
Db 1938 GCCACCGAGCGCTCCGGCAACACCCAGTGGCGGACAGTGTCCACCGTGGACTCGGTAGCC 1879
QY 72 -----ArgThrIlePheThrGlnGlyAlaPheGluPro 82
Db 1878 CCGGCCACCGCAGTGATCAACCCCGACGACCGCACCGTCTACGCGGACCGCGGAGCGG 1819
QY 83 GlyAsnSerValThrAspLeuAlaIleGlyGlyGlyPhePheGlnValThrLeuGlu 102
Db 1818 GCGACGACGCTGACCTCAGCGATGGCAACCGCACCGGATCGGCGGAGGTCTACCGCGGAC 1759
QY 103 AspLysValHisTyrThrArgAlaGlyAsnPheArgPheThr-----Gln 117
Db 1758 GGC-----AGCGCAACTGGAGCTTACCCCGTCCACCGCGCTGGCGG 1717
QY 118 AspGlyPheLeu-----AsnAspProSerGlyPheThrLeuMetGlySer 132
Db 1716 GATGGAACCGTGGTCAACGCGCACCGGATCCGCGCGGCAACACC-----1669
QY 133 ArgIleSerAsnAsnProAsnIleLysGlyLeuThrLeuGluProIleGlnLeuAspPhe 152

Db 1668 -----AGCGCGCAGGCGAGCACCACCGTCGATGGCGCGCG----- 1630
QY 153 AsnAspProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValVal 172
Db 1629 ACCACCGCAGCGTCACTGAGC-----AACGGCAGCGCTCAGCGGCGACTGCG 1579
QY 173 AsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeu 192
Db 1578 GAACCGCGCAGCAGCGTGTCTCCAGCAGCAACGCGCAATCCGATCCCGAGGTACC 1519
QY 193 GluSerTrpLysGlyAsnGly-----ThrProPheLysSerThrSerAsnTyrSerTyr 210
Db 1518 GCGCAGCGCAGCGGCAACTGGACCTACACCGCGCTCCACCGCGATCGCCAAACGCGACCGTG 1459
QY 211 AlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHis-----AspIleThr 227
Db 1458 GTCAACGTGGTG---GCCAGCAGCGCGCGCGCATAGCAGCGCGCGCGCGCGCGTACC 1402
QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
Db 1401 GTG---GACTCGCAAGCGCGCGCGCTCCG-----GTG 1372
QY 248 MetAsnProSerGluAspGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMet 267
Db 1371 GTCAACCGCGAGC---AACGGCAGCAGCTCAGCGCGCAGCGCGCGCGCGCTTACCCTG 1315
QY 268 -----SerGlyThr 270
Db 1314 AGCCTGACCGCAGCGCAACCGCAACCGGATTCGCCAGGTCCACCGCGCAGCGAGTGGCAAC 1255
QY 271 MetThrPheSer-----SerAsnGlyGluLeuLysAsnMetThrAlaPhe 285
Db 1254 TGGAGCTTCACACCGCGCAGCGCGTCCGCGCAGCGCGCGCGCGCGCGCGCGCGCG 1195
QY 286 ThrProThrGlySer-----AlaThrLysAspLeuAsnAlaTyrGlnPro--- 300
Db 1194 GACCGCAGCGCAATACCGAGCGCTCCGCGCAGCACCACCGTGGACTCGTGGCGCGCGCG 1135
QY 301 AlaProLeuValAsn-----GlyLeuProGlnPheSerAla 312
Db 1134 GCGCGCGTGGTCAATCCGAGCAACCGAGTCTCATCAGCGCGCACCGCGCAACCGCGCGCG 1075
QY 313 AsnPheVal-----GlyAlaGly-----IleGlnProLeuThrLeuAspPheGly 327
Db 1074 ACCGTGACCTTCACCGATGGAGCGGCAATCCGATCCGCGAGGTCCCGCGCGAC----- 1021
QY 328 IleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaLalleGlyThr 347
Db 1020 -----GGCAGCGGCAACTGGAGCTTCACCGCGTCCACCGCGCTGGCGGATGAACC 970
QY 348 AspIleGlyLys----- 351
Db 969 GTGTCACGCCACCGCTACGAGCGCGCGCGCAATACCGCGCGCGCGCGCGACACTACC 910
QY 352 -----LeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAla 367
Db 909 GTGACGCCATCCG 850
QY 368 ArgAsnGlySerSerThrArgTyrSerGlnAspGlyTyrProGlnGlyAspLeu 387
Db 849 GGCACCTCGGAACCGCGCAGCGGTGATCTCTCACCGAGCGC-----AACGCGCAATCCG 796
QY 388 ValAspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnVal 407
Db 795 ATC---GCCAGGTCCACCGCGCGCGC-----AGCGCAACTGGACCTACACCGCGTCC 745
QY 408 ValAspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGlu 427
Db 744 ACGCCGATCCCGCAACCGGTACTGTGGTCAACGTGGTGGCGCGCGCGCGCGCGCGCGCG 694
QY 428 GlyAsnAsnHisTyrSerAla-----ThrLeuAspSerGlyGlyProGluPheGlyLeu 445
|||||

Db 693 GGTAAACAGCAGCG 634
QY 446 ProGlyThrSerAsn 450
Db 633 ATCAACCCGAGCAAC 619
RESULT 34
US-08-436-748-2
; Sequence 2, Application US/08436748
; Patent No. 5827654
; GENERAL INFORMATION:
; APPLICANT: CHAN, VOON LOONG
; APPLICANT: LOUIE, HELENA
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1P7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,748
; FILING DATE: 05-AUG-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I
; REGISTRATION NUMBER: 24,973
; REFERENCES/DOCKET NUMBER: 1038-428 MIS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 810 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-436-748-2
Alignment Scores:
Pred. No.: 3,58e-07 Length: 810
Score: 155.00 Matches: 46
Percent Similarity: 49.33% Conservative: 28
Best Local Similarity: 30.67% Mismatches: 64
Query Match: 5.99% Indels: 12
DB: 1 Gaps: 5
US-10-009-823A-1 (1-502) x US-08-436-748-2 (1-810)
QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 1 ATGCAAAATGATATTAATCAAGCAACTGGCGGAATGTAATCAGTTTAATAAATTTGAT 60
QY 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal 40
Db 61 GTGATTACTAATAATCTTGCAATATCAATCAAGTGAATATAAAGAGATGATGTGTT 120
QY 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlalleGlySer--- 53
Db 121 ATTCCAGATTTTAAAGGATTTTAAAGAAACTCAGGATGAGTTGCTATAGAAATCAC 180
QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly---AlaGlnValGlySer 70
Db 181 ACAAGAGATCACTCGTTTGTGTAATAACTACTACAATAGATGAATCCCAAGTTTCTCAA 240

QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
Db 241 GAATATACGATTTAGCTAGCTTCTTTAAAGGCCACAAACAATCTTTGGATTGGCA 300
QY 91 IleGlyGlyLysGlyPheGlnVal-----ThrLeuGluAspLysValHisTyThr 108
Db 301 ATGACTAGAGAAGATGCTTTTATTTGGTTACAGCAACCAAGATGGAGAGTAAGATTAA 360
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 361 AAAGATGGAAATTTCACTTGATGATGAGGTTATTTGGTAAATAAGCAAGGATACAAG 420
QY 129 LeuMetGlySerArgIleSerAsnAsnPro 138
Db 421 GTATTAAAGTAGTATTATTTAATAATCCT 450

RESULT 35
US-08-483-857-2
; Sequence 2, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 810 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-483-857-2

Alignment Scores:
Pred. No.: 3,598-07 Length: 810
Score: 155.00 Matches: 46
Percent Similarity: 49.33% Conservative: 28
Best Local Similarity: 30.67% Mismatches: 64
Query Match: 5.99% Indels: 12
DB: 3 Gaps: 5

US-10-009-823A-1 (1-502) x US-08-483-857-2 (1-810)

QY 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 1 ATGCAAAATGATATTATCAAGCAACTCGCGGAATGTAACCTAGTTTAAATAACTTGA 60
QY 21 ThrValSerAsnAsnIleAlaAsnThrIleGlyTyLysGlnGlnValVal 40

Db 61 GTGATTACTATAATCTTCCCATATCAATCAAGTGGATATAAAAGAGATGATGGTT 120
QY 41 -----PheGlnAspLeuPheSerGln-----AspLeuAlaIleGlySer--- 53
Db 121 ATTGCAGATTTTAAAGGATTTTAAAGAAACTCAGGATGAGTTGCCTATAGAAAATCAC 180
QY 54 -----ThrGlySerGlnGlyProAsnGlnAlaGlyMetGly---AlaGlnValGlySer 70
Db 181 ACAAGAGATGATCTCGTTTGTAAATCACTCAATAGATGAATCCCAAGTTCTCA 240
QY 71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
Db 241 GAATATACGATTTTAGCTAGGTTCTTTAAAGGCCACAAACAATCTTTGGATTGGCA 300
QY 91 IleGlyGlyLysGlyPheGlnVal-----ThrLeuGluAspLysValHisTyThr 108
Db 301 ATGACTAGAGAAGATGCTTTTATTTGGTTACAGCAACCAAGATGGAGAGTAAGATTAA 360
QY 109 ArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThr 128
Db 361 AAAGATGGAAATTTCACTTGATGATGAGGTTATTTGGTAAATAAGCAAGGATACAAG 420
QY 129 LeuMetGlySerArgIleSerAsnAsnPro 138
Db 421 GTATTAAAGTAGTATTATTTAATAATCCT 450

RESULT 36
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1,11 Length: 4403765
Score: 151.50 Matches: 119
Percent Similarity: 31.12% Conservative: 64
Best Local Similarity: 20.24% Mismatches: 220
Query Match: 5.86% Indels: 185
DB: 3 Gaps: 24

US-10-009-823A-1 (1-502) x US-09-103-840A-2 (1-4403765)

QY 3 GlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22
Db 368448 GGCAATATCGCATCGGCTCACCGGC-----ACCGTTCAGATCGGTTCCGCGAGCTTC 368395
QY 23 SerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyLysGlnGlnValValPheGln 42
Db 368394 AACTCGGCGGACCCACACATCGGCTTGTCACTCGGTCGACGGAACCGTAGGATCTTC 368335
QY 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGly----- 55
Db 368334 AACTCGGCGACCGCGCAACGTTGGCGATCGGAAACACCGGCACCGCAAACTTCGCGATCGCA 368275

QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
DB 368274 AACTCGGGGAGCTTCAACACCGGCTCTGGG---AACACGGGAGCAGCAACACG--- 368224
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys--- 94
DB 368223 ---GGCCTGTTCACACCGGGCAACGTCACACCGGCTCTGGGCAACACCGGAGCATC 368170
QY 95 ---GlyPheGlnVal--- 99
DB 368169 AACACGGGAGCATCAACACCGGAGCTTCAACATCGGAGCAGCAACATACCGGAGCTTC 368110
QY 100 ThrLeuGluAsp--- 108
DB 368109 AACCTCGGCGATCAACACCGGAGCTTCAACTCGGCTGACTACACACCGGCTACTTC 368050
QY 109 ArgAlaGlyAsnPheArgPhe--- 115
DB 368049 AACCGGGGTGACTACACACCGGCTGTGCCAACACCGGCAACGTCACACCGGCGCTTC 367990
QY 116 ---ThrGlnAspGlyPhe--- 120
DB 367989 ATCTCGGCAATTACAGCAACCGGCTTCTCTGGGAGGTGACTACCGGGGTGATTGGC 367930
QY 121 ---LeuAsnAspProSer 125
DB 367929 CTTTCCACACGATCACCATTCCTCGGAATCCCTACCGCTACGAGCTTGAGTGTTCACATC 367870
QY 126 GlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsn--- 139
DB 367869 GACATACCATCACCGGACCGTGTGCTGCCACCGCAACACGATTCACCATTCCTCCCGT 367810
QY 140 ---IleLysLysGluThrLeuGluPro 147
DB 367809 TTCCAGATACGAGTCTTCTGCTGCTCGGCGGTGCTTGTCAACGAGATGATCGGCGCC 367750
QY 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
DB 367749 ATCAGATCGATGTCATCAAGTCATCGCATCGCATCGCATCGCATCGCATCGCATCGC 367690
QY 168 LeuAsnAlaValAlaValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
DB 367689 ATGGTTGGCACCGGCGCTTCGGCCCGATCCCATCGCATCGCATCGCATCGCATCGC 367630
QY 188 TyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheIleSerThrSerAsn 207
DB 367629 GGTTC---GGCAACTCGACC--- 367612
QY 208 TyrSerTyAlaGlnProMetArgValTyAspGlnGlnGlyAsnSerHisAspIleThr 227
DB 367611 ---ACCGGCGCTGCTCGGTTCTTCCACACCGGCGCGCATGATCGGC 367561
QY 228 ValTyPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyLeuValAla 247
DB 367560 TTCGGGAATTCGGCGCGGCAACATGTCGGGCTCGCGGAACCTTC--- 367516
QY 248 MetAsnProSerGluAspGlySerAlaAlaSerGly---ThrAspSerAlaGlyLeuLeu 266
DB 367515 ---GGCGTGGCAATTCGGGCTCTTAAACCGCGCGCTTGGC 367474
QY 267 MetSerGlyThrMetThrPheSerSer---AsnGlyGluLeuLysAsnMetThrAlaPhe 285
DB 367473 AATTCGGGCTTACTGAATTCGGGCGCTGTCAGTCTGGGCTTGGCAACCTGGGCAACACC 367414
QY 286 ThrProThrGlySerAlaThrLysAlaLeuAsnAlaTrpGlnProAlaProLeuValAsn 305
DB 367413 ATCTCGGGCTCTACACACGACGACGCTGACCTCGGACGCGCGCTTC---TTCCGCTCG 367357
QY 306 GlyLeuProGlnPheSerAlaAsnPheValGlyValGlyIleGln---ProLeu 322
DB 367356 GGCATCGCAAAACATCGGCGGCAACCTGCGCGGCTGTTCCTCGACACACCGGCAACCTG 367297

QY 323 ThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342
DB 367296 ACCTGAGCTTCGGCTCGCAACACGAGCGGCTCAACCGCGGC--- 367252
QY 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362
DB 367251 ---ATCGGAACCTGGGAGCGTCAACATCGGCTTCGTTAATACC 367210
QY 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTy--- 377
DB 367209 GGCAGCTCAACCTGGGATCGGCACTCGGACCTCGGCGACCTCAACTCGGCGGTCAACATC 367150
QY 378 ---SerGlnAspGlyTyProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
DB 367149 GCGGGTAACACATCGGATCGCAACACCGGATCTTCGATATCGGCTTGGCGAACCTG 367090
QY 396 GlyLysLeuGlnGlyLysTySerAsn---SerGlnValValAspPhe--- 410
DB 367089 GGCAGCTACACATCGGCTTGGCAATCTGGGCGACGACCACTGGGCTTGGCAACGCC 367030
QY 411 ---TyAsnIleProLeuAlaArgPheThrSerGluAsp--- 422
DB 367029 GGCAGCTACACATCGGCTTGGCAATCTGGGCGACGACCACTGGGCTTGGCAACGCC 366970
QY 423 ---GlyLeuArgArgGluGlyAsnAsnHisTySer--- 433
DB 366969 GGCAGCTACACATCGGCTTGGCAATCTGGGCGACGACCACTGGGCTTGGGCTCACC 366910
QY 434 ---AlaThrLeuAspSerGlyGlyProGluPheGlyLeu--- 445
DB 366909 GGCACCGCCAGATCGGATCGGAGCTCACTCGGCGACGACCAACATCGGCTGTTTC 366850
QY 446 ---ProGlyThrSerAsnTyGly 452
DB 366849 AACTCGGCGAGGAAACATCGG 366826

RESULT 37
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Alignment Scores:
Pred. No.: 1.11 Length: 4411529
Score: 151.50 Matches: 119
Percent Similarity: 31.12% Conservative: 64
Best Local Similarity: 20.24% Mismatches: 220
Query Match: 5.86% Indels: 185
DB: 3 Gaps: 24

US-10-009-823A-1 (1-502) x US-09-103-840A-1 (1-4411529)

QY 3 GlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrVal 22
DB 368391 GGCAATATCGGATCGGCTCACCAGC-----ACCGTCAAGATCGGCTTCCGCGAGCTTC 368338

QY 23 SerAsnAsnIleAlaAsnAlaThrIleGlyTyrLysGlnGlnInValValPheGln 42
Db 368337 AACTCGGAGCCACCAATCGCTTGTCAACTCCGGTGTACGGAACAGTAGATTCTTC 368278
QY 43 AspLeuPheSerGlnAspLeuAlaIleGlySerThrGly----- 55
Db 368277 AACTCGGAGCCACCAATCGCTTGTCAACTCCGGTGTACGGAACAGTAGATTCTTC 368218
QY 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePhe 75
Db 368217 AACTCGGAGCCACCAATCGCTTGTCAACTCCGGTGTACGGAACAGTAGATTCTTC 368167
QY 76 ThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGlyGlyLys-- 94
Db 368166 -----GGCTGTTTCAACCGCGGCAACGTCACACCGCGCTCGGCAACACCGCGAGCATC 368113
QY 95 -----GlyPheGlnVal----- 99
Db 368112 AACACGGGAGCATCAACACCGCGAGCTTCAACTCGGAGCACCACCAATACCGGAGCTTC 368053
QY 100 ThrLeuGluAsp-----LysValHisTyrThr 108
Db 368052 AACTCGGAGCCACCAATCGCTTGTCAACTCCGGTGTACGGAACAGTAGATTCTTC 367993
QY 109 ArgAlaGlyAsnPheArgPhe----- 115
Db 367992 AACCGGGTGTACTACACACCGGCTGTGCGCAACAGCGGCAACGTCACACCGCGCGTTC 367933
QY 116 -----ThrGlnAspGlyPhe----- 120
Db 367932 ATCTCCGGAATTACAGCAACGGTCTTCTTGGCGAGGTGACTACACAGGGGTTGATTGGC 367873
QY 121 -----LeuAsnAspProSer 125
Db 367872 CTTTCCACACAGATCACCATTCCCGAATCCCTTACCGCTACGAGTGTGAGTGTTCATC 367813
QY 126 GlyPheThrLeuMetGlySerArgIleSerAsnProAsn----- 139
Db 367812 GACATACCATCACCAGGACCGTGTGCGCACACCGCAACAGTTCACCATTCCTCCGGT 367753
QY 140 -----IleLysLysGluThrLeuGluPro 147
Db 367752 TTCCAGATACAGTCTTCTGTCCTGCGCGGTGTCTGTCAACAGATGATCGGCGCC 367693
QY 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSerProAlaLysThrSerThrAla 167
Db 367692 ATCAGATCATGATCAATCAATGATCATCGCATCGCTTCCGCTTCAAGCAACCATCAGC 367633
QY 168 LeuAsnAlaValValAsnLeuGlyAspSerThrAspLysThrGlnSerGluAlaAsnPro 187
Db 367632 ATGGTTGGACCGCGCGGTTCGCGCCGATCCCATCGGCATCAGCATCGGTGTACCCCG 367573
QY 188 TyrPheAlaLeuLeuGluSerTriLysGlyAsnGlyThrProPheSerThrSerAsn 207
Db 367572 GGTTC-----GGCAACTCGACC----- 367555
QY 208 TyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAspIleThr 227
Db 367554 -----ACGCGCGCTCGTGGTTCCTCCACACCGCGCGCCCATGTATCGGC 367504
QY 228 ValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuValAla 247
Db 367503 TTCCGGAACCTCGCGCGGCAACATGTCGGCTCGGGAATTC----- 367459
QY 248 MetAsnProSerGluAspGlySerAlaAlaSerGly-----ThrAspSerAlaGlyLeuLeu 266
Db 367458 -----GGCGTGGCAATTCGGGCTTCTTTTAAACCGCGCGCTTGGGC 367417
QY 267 MetSerGlyThrMetThrPheSerSer-----AsnGlyGluLeuLysAsnMetThrAlaPhe 285
Db 367416 AATTCGGCGCTTACTGAATTCGGCGCGCTGCGAGTCGGGTCTGGCAACCTGGGCAACACC 367357
QY 286 ThrProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsn 305

Db 367356 ATCTCGGCGGTCTCAACACAGCAGCAGCTCGACCTCGCGAGCGCCGCTTC 367300
QY 306 GlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGln-----ProLeu 322
Db 367299 GGCATCGCAACATCGCGCGCAACCTCGCGGCTGTTCCTCGCAACACCGCAACCTG 367240
QY 323 ThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAla 342
Db 367239 AGCTGAACTTCGGCGTGCMAACAGCGCGGCTCAACCGGCGC----- 367195
QY 343 AlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSer 362
Db 367194 -----ATCGGAACCTCGCGAGCGTCAACATCGGCTTCGTTAATACC 367153
QY 363 GlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArgTyr----- 377
Db 367152 GCGGACTCAACCTGGGATCGGCAACTCGGCACTCGGCACTCAACTTCGGCGGGTCAACATC 367093
QY 378 -----SerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
Db 367092 GCGGTAAACATCGGATCGGCAACACCGGATCTTCGATATCGGCTTGGCGAACCTG 367033
QY 396 GlyLysLeuGlnGlyLysTyrSerAsn-----SerGlnValValAspPhe----- 410
Db 367032 GCGACTTAAACATCGGCTTGGCAATCTCGGCACTCGGCACTCGGCTTGGCAACGCC 366973
QY 411 -----TyrAsnIleProLeuAlaArgPheThrSerGluAsp----- 422
Db 366972 GCGAGTAAACATCGGCTTGGCAACTCGGCACTCGGCACTCGGCTTGGCAACGCC 366913
QY 423 -----GlyLeuArgGluGluGlyAsnAsnHisTyrSer----- 433
Db 366912 GCGAGTAAACATCGGCTTGGCAATCTCGGCACTCGGCACTCGGCTTGGCGGTCAACC 366853
QY 434 -----AlaThrLeuAspSerGlyGlyProGluPheGlyLeu----- 445
Db 366852 GCGAACGCCAGATCGGATCGGCACTCGGCACTCGGCACTCGGCACTCGGCTTGGC 366793
QY 446 ---ProGlyThrSerAsnTyrGly 452
Db 366792 AACTCGGCGAGCGGAACATCGGG 366769

RESULT 38
US-09-206-942-64
; Sequence 64, Application US/09206942
; Patent No. 6432569
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 64
; LENGTH: 3543
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-09-206-942-64

Alignment Scores:
Pred. No.: 1.57e-05 Length: 3543
Score: 149.50 Matches: 113
Percent Similarity: 37.09% Conservative: 91
Best Local Similarity: 20.55% Mismatches: 220
Query Match: 5.78% Indels: 127
Gaps: 21

US-10-009-823A-1 (1-502) x US-09-206-942-64 (1-3543)

QY 30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerClnAspLeu 49
 Db 1816 AATATTTCAGGCTTCACAAACCAAGATTGTAGCTAAGAT-----AGTAGTAATTTA 1869
 QY 50 AlalleGlySerThrGlySerGlnGlnProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
 Db 1870 ACTATTGGTAATAGTATGATGATAGCGCAATACT-----1902
 QY 70 SerValArgThrIlePheThrGlnGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
 Db 1903 ACGGCTAAACAGTA-----ACTTTTAACAATGTTTAAAGATTCA 1941
 QY 90 AlalleGlyGlyLysGlyPheGlnValThrLeuGluAspLysValHisThrArg 109
 Db 1942 AAAATCTCTGCTGACGGT--CACAAAGGTGACACTAAATAGCAAGTGAAA--ACACTT 1995
 QY 110 AlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeu 129
 Db 1986 AGTGATAATGATAACAAACACTGAAGTGGCGAGTGACACAAAT--ACCGGTTTAACTATT 2052
 QY 130 MetGlySerArgIleSerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGln 149
 Db 2053 ACTGCAAAAGATGTAGAGTAAACACAAATATTACT-----2088
 QY 150 LeuAspPheAsnAspProThrValAlaLysSerProAla-----Lys 163
 Db 2089 -----TCTCAAAACAGTGAACGTCTCTCGGCAAAATGAGGAGTATACCACATAA 2139
 QY 164 ThrSerThrAlaLeuAsnAlaVal-----ValAsnLeuGlyAspSerThrAspLys 180
 Db 2140 ACAGGTACACCAATTAATGCAACCGCGGTACGTGAGAGTAAACCGCTCATACAGGCAGT 2199
 QY 181 ThrGln-SerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGly-----197
 Db 2200 ATCCAAGCGGAATGTAGTCAAGCTGGCTCTGTGACAAATTTGGCAGCGCGGATACT 2259
 QY 198 -----AsnGlyThrProIleSerThrSerAsn-----207
 Db 2260 CTTCGTGTAGTAATATTTCAGGCAACCGCGGTACTCTT-ACGTCAATATAGCGGTGCATT 2318
 QY 208 -----TyrSerTyrAlaGlnProMetArgValTyrAspGlnI 220
 Db 2319 AACCACTTTGCGAGGCTCTACAAATTAAAGGACCGAGAGTATAACCACTTCAAGTCAATC 2378
 QY 220 nGlyAsn-----SerHisAspI 226
 Db 2379 AGGTAATATCGCGGTAAAAATTTCCGGCAAGACAGATAAAGCTTAAAGCACTAATAGTTT 2438
 QY 226 eThrValTyrPheAspGlyAlaProSerSerThrGlySerLysThrPheGluTyrLeuVa 246
 Db 2439 AACCAACCAAGCAGACTCAAAATTTAGAGCATGTAGGC-----GAGCTAATGT 2489
 QY 246 lAlaMetAsnProSerGluAspGlySerAlaAlaSerGly-----259
 Db 2490 AACCAAGCAAAACAGCAATAATTTGGCGGTACAATTTCTGTGTGGCACAGTAGAAGTTACCGC 2549
 QY 260 -----ThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSe 275
 Db 2550 GACCGAAGGTTTAAACCCACAGCGCTCTACGATTACTGGAAACGAGCGGTGACCA 2609
 QY 275 rAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGly-----SerAl 291
 Db 2610 TTCAAGCAATCAGGTAATATCGCGGCATGATTTCTGTGTGGCAAGTAGAAGTTAGCGC 2669
 QY 291 aThrLysAspLeu-----As 296
 Db 2670 AACCAAGATTTTAACTAAATTCGGTTCAGAGATTAAGCAACGCGCGGAGGTGAA 2729
 QY 296 nAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValG 316

2730 TGTAAACAAGTGCACACAGGTACAAATTGACCGGTACGATTTCCGGTAAATACGTAATGTTAC 2789
 QY 316 YAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAl 336
 Db 2790 AGCAATATCTGCGCATTTAACTGTTGAAGATCGCGCAAAA-----AT 2831
 QY 336 aGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMe 356
 Db 2832 TGATGCGACAGAGGAGCGCGACCTTAATCTGCAACATCGGCAAAATTA-----2880
 QY 356 tProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArgAr 376
 Db 2881 -----ACCACCTAAGCTAGTTTCAAGCATTAATCTTCAAGCTAATAACCAAGTAAACCT 2930
 QY 376 gTyrSerGlnAspGlyTyrProGlnGlyAspLeu-----ValAspValThrIleThrSe 394
 Db 2931 TTCAGCTAAGGATGTAGCATTTGGGGGAATATCAATGCTGCTAATGTAACTGAATAC 2990
 QY 394 rGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePr 414
 Db 2991 TACAGCGCTCTAACTACCGTGAAGGTTCAAGCATTAACGCCAACAGCGGACCTTGCT 3050
 QY 414 oLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAl 434
 Db 3051 TATTAACGCAAAAGACGCTGAGCTAAATGGTGAGGCATCAGGTAAACCATACAGTAGTAA 3110
 QY 434 aThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLe 454
 Db 3111 TCCAACCAACGCAATGCTCCGCGAGCGTAAATCCGCAACACCTCAACAGAGTCAACAT 3170
 QY 454 uSerValAsnGlnLeuGluThrSerAsnValAsp---MetSerArgGluMetValAsnMe 473
 Db 3171 CACTCGGATTTAATCAACAATAATGGATTAAATATCATTTTCAAAAAACGGTATATAACAC 3230
 QY 473 tIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThrMe 493
 Db 3231 CGTACTGTATA--AAAGCGTTAAATTTGAATGTGAATACATTCACCGGTTAGCAAG 3287
 QY 493 tLeuGlnLysAlaLeuGluLysArg 502
 Db 3288 CGTAGATGAAGTAATTTGAAGCGAAACGC 3315

RESULT 39
 US-09-206-942-62
 ; Sequence 62, Application US/09206942
 ; Patent No. 6432669
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomis, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 ; FILE REFERENCE: 1038-861 MIS:jb
 ; CURRENT APPLICATION NUMBER: US/09/206,942
 ; EARLIER FILING DATE: 1998-12-08
 ; EARLIER APPLICATION NUMBER: 09/167,568
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 62
 ; LENGTH: 3568
 ; TYPE: DNA
 ; ORGANISM: Haemophilus influenzae
 US-09-206-942-62

Alignment Scores:
 Pred. No.: 1.59e-05 Length: 3568
 Score: 149.50 Matches: 113
 Percent Similarity: 37.09% Conservative: 91
 Best Local Similarity: 20.55% Mismatches: 220
 Query Match: 5.78% Indels: 127
 DB: 4 Gaps: 21

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:09:07 ; Search time 132 Seconds

(without alignments)
1231.264 Million cell updates/sec

Title: US-10-009-823A-1

Perfect score: 2586

Sequence: 1 MMGSLFIGATGMKTHSTGLG.....NGSKVTTADTLMQLKALELKR 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1364641 seqs, 323758627 residues

Total number of hits satisfying chosen parameters: 1364641

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2586	100.0	502	14	US-10-009-823A-1
2	529.5	20.5	414	15	US-10-009-823A-1
3	527.5	20.4	462	14	US-10-127-032-101
4	506.5	19.6	413	15	US-10-127-032-101
5	485	18.8	718	15	US-10-335-977-4820
6	355.5	13.7	477	14	US-10-009-823A-7
7	351	13.6	462	14	US-10-009-823A-8
8	337.5	13.1	400	14	US-10-009-823A-10
9	333	12.9	401	14	US-10-009-823A-9
10	296	11.4	221	15	US-10-335-977-4818
11	288	11.1	262	15	US-10-335-977-4763
12	288	11.1	271	15	US-10-335-977-4764
13	288	11.1	279	15	US-10-335-977-4765

14	282	10.9	480	14	US-10-009-823A-11	Sequence 11, Appl
15	233	9.0	360	14	US-10-009-823A-12	Sequence 12, Appl
16	201.5	7.8	152	15	US-10-335-977-4840	Sequence 4840, Ap
17	189.5	7.3	464	13	US-10-335-977-4844	Sequence 4844, Ap
18	186	7.2	370	14	US-10-009-823A-13	Sequence 13, Appl
19	180.5	7.0	655	14	US-10-369-493-4713	Sequence 4713, Ap
20	180.5	7.0	655	14	US-10-369-493-4713	Sequence 4713, Ap
21	173.5	6.7	1095	14	US-10-193-764-65	Sequence 65, Appl
22	173.5	6.7	1536	13	US-10-092-680-2	Sequence 2, Appl
23	173.5	6.7	1536	14	US-10-193-764-63	Sequence 63, Appl
24	172.5	6.7	3705	15	US-10-282-122A-77944	Sequence 77944, A
25	166	6.4	91	15	US-10-335-977-4817	Sequence 4817, Ap
26	166	6.4	2468	14	US-10-246-330-4	Sequence 4, Appl
27	166	6.4	2468	15	US-10-282-122A-66335	Sequence 66335, A
28	165.5	6.4	124	15	US-10-335-977-4839	Sequence 4839, Ap
29	164	6.3	343	15	US-10-335-977-4819	Sequence 4819, Ap
30	162	6.3	1954	15	US-10-147-299A-4	Sequence 4, Appl
31	160	6.2	1230	9	US-09-881-752A-150	Sequence 150, App
32	160	6.2	1230	15	US-10-282-122A-58893	Sequence 58893, A
33	156.5	6.1	3073	15	US-10-282-122A-49147	Sequence 49147, A
34	153.5	5.9	679	14	US-10-369-493-13825	Sequence 13825, A
35	151.5	5.9	269	15	US-10-335-977-4779	Sequence 4779, Ap
36	151.5	5.9	2204	15	US-10-282-122A-64364	Sequence 64364, A
37	149.5	5.8	1180	14	US-10-193-764-61	Sequence 61, Appl
38	149.5	5.8	1188	14	US-10-193-764-59	Sequence 59, Appl
39	148.5	5.7	273	15	US-10-335-977-4780	Sequence 4780, Ap
40	147	5.7	1649	14	US-10-369-493-18460	Sequence 18460, A
41	146.5	5.7	486	14	US-10-316-175-12	Sequence 12, Appl
42	143	5.5	1331	15	US-10-282-122A-47930	Sequence 47930, A
43	140.5	5.4	529	14	US-10-316-175-15	Sequence 15, Appl
44	139.5	5.4	1797	14	US-10-369-493-5176	Sequence 5176, Ap
45	139.5	5.4	1805	14	US-10-369-493-5177	Sequence 5177, Ap

ALIGNMENTS

RESULT 1

US-10-009-823A-1
; Sequence 1, Application US/10009823A
; Publication NO. US20030157120A1
; GENERAL INFORMATION:
; APPLICANT: Panaccio, Michael
; APPLICANT: Rosey, Everett Lee
; APPLICANT: Sinistaj, Meri
; APPLICANT: Hasse, Detlef
; APPLICANT: Parsons, Jim
; APPLICANT: Ankenbauer, Robert G.
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE
; FILE REFERENCE: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES
; FILE REFERENCE: DAVI150.001APC
; CURRENT APPLICATION NUMBER: US/10/009,823A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/AU00/00437
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/133,973
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-10-009-823A-1

Query Match 100.0%; Score 2586; DB 14; Length 502;
Best Local Similarity 100.0%; Pred. No. 1.6e-202; Indels 0; Gaps 0;
Matches 502; Conservative 0; Mismatches 0

QY 1 MMGSLFIGATGMKTHSTGLTGVNNIANANTIGYKQQQVWFQDLFSQDLAIGTSGQPN 60
DB 1 MMGSLFIGATGMKTHSTGLTGVNNIANANTIGYKQQQVWFQDLFSQDLAIGTSGQPN 60

```
QY 61 QAGMAQVGSVRTIIFQGAPEFGNSVTDLAIGKGFFQVLTLEDKVHYTRAGNFRFTQDGF 120
Db 61 QAGMAQVGSVRTIIFQGAPEFGNSVTDLAIGKGFFQVLTLEDKVHYTRAGNFRFTQDGF 120
QY 121 LNDPSGFTLMSRISNNPNIKKTELPIQLDFNDPTVAKSPAKTSTALNAVNLGDSYDK 180
Db 121 LNDPSGFTLMSRISNNPNIKKTELPIQLDFNDPTVAKSPAKTSTALNAVNLGDSYDK 180
QY 181 TQSEANPYFALLESKWKGNGTPTPISSTNSYSAQPMRVYDQGNSHDITVYFDGAPSSGSK 240
Db 181 TQSEANPYFALLESKWKGNGTPTPISSTNSYSAQPMRVYDQGNSHDITVYFDGAPSSGSK 240
QY 241 TFEYLIVAMPSPEDGSAAGSTDSAGLLMSGTTFSSNGELKNMTAFTPTGSATKDLNAWOP 300
Db 241 TFEYLIVAMPSPEDGSAAGSTDSAGLLMSGTTFSSNGELKNMTAFTPTGSATKDLNAWOP 300
QY 301 APLVNLGPOFSANFVGAGIQPLTLDFGIKSQONMWAGAPASAAAIGTIDIGKLPSPMPTQT 360
Db 301 APLVNLGPOFSANFVGAGIQPLTLDFGIKSQONMWAGAPASAAAIGTIDIGKLPSPMPTQT 360
QY 361 SSGNSTANGSSSTRYSQDGHPOQDLVDVITITSEGLKQKYSNSQVVDVFNIPARPTS 420
Db 361 SSGNSTANGSSSTRYSQDGHPOQDLVDVITITSEGLKQKYSNSQVVDVFNIPARPTS 420
QY 421 EDGLRREGNNHYSATLDSGGPEFGPLPTSNYKGLSVNOLETSNVDMGSMVMNIIIOQGF 480
Db 421 EDGLRREGNNHYSATLDSGGPEFGPLPTSNYKGLSVNOLETSNVDMGSMVMNIIIOQGF 480
QY 481 QMSKSVTTADTLMLOKALELKR 502
Db 481 QMSKSVTTADTLMLOKALELKR 502

RESULT 2
US-10-282-122A-48141
; Sequence 48141, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
```

```
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48141
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
; US-10-282-122A-48141

Query Match 20.5%; Score 529.5; DB 15; Length 414;
Best Local Similarity 29.8%; Pred. No. 1.4e-34;
Matches 149; Conservative 72; Mismatches 182; Indels 97; Gaps 13;

QY 8 GATGKTHGTGLTGVNNIANANTIGYKQQQVVFQDLFSDLAIGSTGSGGNQAGMAQ 67
Db 6 GLSGLAGASSNLDVIGNNIANANTVGFQGRANFADMYANSVATSVN-----TQIGIGTR 60
QY 68 VGSVRTIIFQGAPEFGNSVTDLAIGKGFFQVLTLEDKVHYTRAGNFRFTQDGF 127
Db 61 LASVQNFQGTINSTKSLDVAINGNFFQFQSSNGVTYSRDTGTHRDKNGAIYDAQGR 120
QY 128 TLMGSRISNNPNIKKTELPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTSEANP 187
Db 121 NLMGYAAGAGGVINTAQTVPLOA-----PTNNIAPRATS-KITGQFNL-NAQDKVPK 171
QY 188 YFALLESKWKGNGTPTPISSTNSYSAQPMRVYDQGNSHDITVYFDGAPSSGSKTFEYL 245
Db 172 -----TPFNATDNTTNTNSISQVDTTGGSQQVTFYF--AKSAGT----- 211
QY 246 VAMPSPEDGSAAGSTDSAGLLMSGTTFSSNGELKNMTAFTPTGSATKDLNAWOPAPLVN 305
Db 212 -----WLAYGVQGGTPTNLTGTVTFDASGRISSTTS-AATG-----QPTP--- 250
QY 306 GLPQFSANFV-----GAGIQPLTLDFGIKSQONMWAGAPASAAAIGTIDIGKLPSPMPTQS 361
Db 251 SLGQFAFSIPTNTGGANPQNLTLDL----- 275
QY 362 SGNSTANGSSSTRYSQDGHPOQDLVDVITITSEGLKQKYSNSQVVDVFNIPARPTS 421
Db 276 -GGTQYGGKQGVNLAQDGFASGTLTTFSTGTDGKLAGNSNGQSAVLGLIALANFNP 334
QY 422 DGLRREGNNHYSATLDSGGPEFGPLPTSNYKGLSVNOLETSNVDMGSMVMNIIIOQGF 481
Db 335 NGLVNIIGNQYAEATAASGVPOIAPFGSTHGTGLOGSALENSNVNLTTLVNLITAQRYQ 394
QY 482 MNSKSVTTADTLMLOKALELKR 501
Db 395 ANAQTIKQOAVDQTLNLR 414

RESULT 3
US-10-127-032-101
; Sequence 101, Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Banger, M. Gita
; APPLICANT: Lory, Stephen
; APPLICANT: Greenberg, Everett Peter
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
; FILE REFERENCE: UIZ-070CP
; CURRENT APPLICATION NUMBER: US/10/127,032
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,190
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/344,142
; PRIOR FILING DATE: 2001-10-24
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 462
```



```

CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)742-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 4820:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...718
SEQUENCE DESCRIPTION: SEQ ID NO: 4820:
US-10-335-977-4820

Query Match 18.8%; Score 485; DB 15; Length 718;
Best Local Similarity 25.0%; Pred. No. 1.3e-30;
Matches 181; Conservative 94; Mismatches 222; Indels 226; Gaps 22;

QY 1 MMSLFGATGKTHSTGLGTVSNINIANANTIGYKQOVVFDLFSQDLAIGS----TGS 56
DB 1 MLRLWSGVNGMQAHLALDIESNNIANVNTGFKYSRASFDVMDLSQVKLIATAPYKNGL 60

QY 57 QGPN--QAGMGAQVSVRTITQGAPEFGNSVTDLAIGKGFQVLEDKV--HYTRAGN 112
DB 61 AGQDNFVGLGVGDVATTKIFSQGNQNTDVKTDLAIGQDGFIIISPDRGITRNFTRDGE 120

QY 113 PRFTQDGLNDPSGFTLM-----GSRISNNPNIKKTELEPIQLDFNDPTVAKSPAK 163
DB 121 FLFDSQGLVTTGGLVGVQVRNGSDTGKSGDTDALKVDNTGPLENIRIDPGMW-WPAR 179

QY 164 TST--ALNAVNLGDDSTKTCSEANFPALLEBK--GNGTPPISSTNSYSYAQPMR---- 215
DB 180 ASNRISMRLNLAGRHADQTA-----VFALDLSAKTPSDGINFVYDSGTNLAQVAEDMGS 235

QY 216 VYDQGN-----SHDITVYFDGA-----PS-----STGSKTFEY 244
DB 236 LYNEGDGALLNENQGLVWYKSAKMKVKDILPAENSTLELNGVKISFTNDSAVRTSSL 295

QY 245 LVANPSEDGSAAG-----TDSAGLMSGTMTFSSNGELKN----- 281
DB 296 VAAKNAINAVKSGTIGIBAYLDGKQLRENTNELDGDEKLNIIVVTQAGTCAFANFLDGDGK 355

QY 282 -MTAF-----TPTGSATKDLNAWAPALVNGLPQFSANFV-----GAGIQPLTL 324
DB 356 DVTAFKYSYTHSISPNADICQFTEDRLALIGHDANIWKDPSLADNYQDSASIGVTIN 415

QY 325 DFG-----IKSQQNMW-----AGAPASAA 343

CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)742-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 4820:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...718
SEQUENCE DESCRIPTION: SEQ ID NO: 4820:
US-10-335-977-4820

Query Match 13.7%; Score 355.5; DB 14; Length 477;
Best Local Similarity 23.7%; Pred. No. 2.8e-20;
Matches 142; Conservative 68; Mismatches 171; Indels 217; Gaps 21;

QY 1 MMSLFGATGKTH-----STGLGTVSNINIANAN 30
DB 1 MMSLFGVSGMNHQGVNPKVGLGVWVASGVNPKVGLGVWVASVTRMDVIGNNVANVN 60

QY 31 TIGYKQOQ-----VVFDLFSQDLAIGSTGSGGPNQAGMGAQVG 69
DB 61 TTGFRGRDITVHTQALQTTGINTDIAIVNFQLISQQL-----SGASRNE-----EVG 111

QY 70 SVRTITQGAPEFGNSVTDLAIGKGFQVLEDKVHYTRAGNFRFTQDGLNDP--SGFT 128
DB 112 Q-----GNGFFILKQGEKSPYTTAGAFGVDRDGTVLNPANGAC 149

QY 129 LMSGSRISNNPNIKKTELEPIQ-LDFNDPTVAKSPAKTSTALNA-----VNLGDSSTKQTS 183
DB 150 NLDKRLMRVQGMMAEDIEGQOIINTSDQPELPGCANQADILRSTEDLIITPIGQKID---- 205

QY 184 EANPYFALLESWKNGTTPPISTNSYSYAQPMRVYDQGNSHDITVYFDGAPSGTG----- 238

```

Db 206 -----AKATTDA-----YTWATDFNYVDTFGQHKLQWVFPVPGTNNQWLAT 249
Qy 239 --SKTPEYLWAMPSPDSGSAAS-----GTDAG-LLMSGT-----MTFSSNGELKNMT 283
Db 250 VVTDAGNVAPNVDENQAGTETRVIGITTDAGQVLVQATENTFVDFDNYGHAS-- 307
Qy 284 AFTPTGSATKDLNAPWAPLVNGLPFSANFVGA-----GIQPIJLDFGKSKQONWAGAP 339
Db 308 -----SYNVGANPDGGAAPRHTFENINDQSGIITGY 340
Qy 340 ASAAA---IGT-----DIGKLPMSMPIQTSSGNSTARNSSSTRYSQDGY 384
Db 341 SNAGSEGEIGTRNITQPAEREIQQL-ALAGFANQGGLEKA--GESITTKAYQDGYAM 397
Qy 385 GDLVDVITITSEGKLOGKYSNSQVVDYFNIPLARFTSEDLRREGNNHYSATLDSGGPEFG 444
Db 398 GYLENFKIT-----YIQSNNSGIANIT 419
Qy 445 LPGTSNYKLSVNQLETSNVDSREVMNMIIIOGFQMSKSVTTADTMLQKALELKR 502
Db 420 VSGVMGKGKLIAGTLENSVDLTDQFTDMIITQGFQAGAKTIQTSDTMLTVLSLKR 477

RESULT 7

US-10-009-823A-8
; Sequence 8, Application US/10009823A
; Publication No. US20030157120A1
; GENERAL INFORMATION:
; APPLICANT: Panaccio, Michael
; APPLICANT: Rosey, Everett Lee
; APPLICANT: Sinistaj, Meri
; APPLICANT: Hasse, Detlef
; APPLICANT: Parsons, Jim
; APPLICANT: Akenbauer, Robert G.
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE
; FILE REFERENCE: DAVI150.001APC
; CURRENT APPLICATION NUMBER: US/10/009,823A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/AU00/00437
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/133,973
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Treponema pallidum
US-10-009-823A-8

Query Match 13.6%; Score 351; DB 14; Length 462;
Best Local Similarity 23.0%; Pred. No. 6.2e-20;
Matches 135; Conservative 67; Mismatches 176; Indels 208; Gaps 17;
Qy 1 MMGSLFTGATGMKTH-----STGLGTVSNNIANANTIGYKQQQVWFQDLF 45
Db 1 MRSFLPSGVSGMNHQGVNPKVGLGLVLIASTRMDVIGNNVANVTGPKR----- 51
Qy 46 SQDLAIGSTGSGPNQAGMGAQVSVRTIETQGAFFPGNSVTDLAI----- 91
Db 52 -----GRDVTHTQALQTTGINTDVSIVNFQDLISQOLSA 88
Qy 92 -----GGKGFQVTLTDKHYTRAGNFRFTQDGLNDPSGFT----- 128
Db 89 ARPNEEVQSGGFPFLVLSGKFTPTFRAGFVDNAGTLVNPANGACNLDKRLMRVQGWMA 148
Qy 129 ---LMGSRISNNPNIKETLEP-IQDLPNDPTVAKSPAKTSTALNAVNLGSDTDKQSE 184
Db 149 QDDVAGERLINS-----SAQPELAADANEADVRS-----TQDLVIPGQKIDAQQT- 195
Qy 185 ANPYFALLESWKNGTTPPISTSNYSYAQPVRVYDQGNSHDITVYFDGAPSGTSKTFEY 244

Db 196 -----STVHYVTWTDFOVDYDFGQQHTLQINFRVPQT----- 228
Qy 245 LVAMKPSDGSAAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATKD-----LNAAW 298
Db 229 -----NNQWQATVVTDAGNVTPAVDPGTEVDQ-----TRVGVGTSDDGTGQVLLLEAA 277
Qy 299 QPAPLVN---GLPQFSANFVGAIGI---QPLTLDFGKSKQONWAGAPASAAAIG----- 346
Db 278 ANTFIVNFDNFHGLASSYDVVGANPDAGQVTRHAFITLNDOSGVITGVYSGVSLGIGT 337
Qy 347 -----TDIGKLPMSMPIQTSSGNSTARNSSSTRYSQDGYPGQDLDVDTITISEG 396
Db 338 ARNITQFAEQDIIQQL-ALAGFANQGGLEKA--GESITTKAYQDGYAMYLENFKIT--- 391
Qy 397 KLOGKYSNSQVVDYFNIPLARFTSEDLRREGNNHYSATLDSGGPEFGPGLPOTSNYKLSV 456
Db 392 -----YVQSNNSGIANISTSGVMGKGKLI 416
Qy 457 NQLETSNVDSREVMNMIIIOGFQMSKSVTTADTMLQKALELKR 502
Db 417 GTLENSVDLTDQFTDMIITQGFQAGAKTIQTSDTMLTVLSLKR 462

RESULT 8

US-10-009-823A-10
; Sequence 10, Application US/10009823A
; Publication No. US20030157120A1
; GENERAL INFORMATION:
; APPLICANT: Panaccio, Michael
; APPLICANT: Rosey, Everett Lee
; APPLICANT: Sinistaj, Meri
; APPLICANT: Hasse, Detlef
; APPLICANT: Parsons, Jim
; APPLICANT: Akenbauer, Robert G.
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE
; FILE REFERENCE: DAVI150.001APC
; CURRENT APPLICATION NUMBER: US/10/009,823A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/AU00/00437
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/133,973
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(400)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-009-823A-10

Query Match 13.1%; Score 337.5; DB 14; Length 400;
Best Local Similarity 25.1%; Pred. No. 6.4e-19;
Matches 128; Conservative 69; Mismatches 165; Indels 147; Gaps 21;
Qy 11 GKMTHSTGLGTVSNNIANANTIGYKQQQVWFQDLFSQDLAIGSTGSGPNQAGMGAQVGS 70
Db 21 GKVAGTLDVIGNNIANSATYGFASGTI-----TQDFTDGTITNG---TGLDVALA- 70
Qy 71 VRTIFTOGAFFPGNSVTDLAIGKGFQFQ-VTLEDKHYTRAGNFRFTQD-----GFLN 122
Db 71 -----SFADMFASQNGFRLVDNSGVSFYSRNGQFKLDENRLNVNMQGIN 116
Qy 123 ----DP-----SGFTLMGSRISNNPNIKETLEPIQDLPNDPTVAKSPAKTSTALNAVNL 174
Db 117 LNSDPLQLQLTGYPATGT-----PPTIQGA-----NFTVTFPSASADSNKPTNI 163
Qy 175 GSDTDKTOSEANPYFALLESWKNGTTPPISTSNYSYAQPVRVYDQGNSHDITVYFDGAP 234

Db 164 --SPNTLMAK-----TITTSKMGSTVTFDSQNAHDMKVYP----- 200
QY 235 SSTGSKTFE-YLVAMNPSDESAAGTDSAGLLMSGTMTSSNGELKNMTAFTPTGSATK 293
Db 201 VKTGTNOKVYFGAMANNQDSSDPSI--AKIATATTFEANGTL-----VDTCG-- 249
QY 294 DLNAWQAPLVNGLPOFSANFVGAGIQPLTDLFGIKSQONWAGAPASABAIGTDIGKLP 353
Db 250 --INGAEPATP--SLSNDGTVVGNYSNEQT--FLNSMQON-----TCQLL 289
QY 354 SMPIQTSSGN--STARNGSSSTRYSQDGYPGQLVDVTTITSEGKLOGKYSNQWDFYN 412
Db 290 GQIVLANFANNEGLASEGDNIVATTQNGYKPGDLVSVQIV----- 330
QY 413 IPLARFTSEDLRREGNNHYSATLDSGPEGLCTSNYKGLSVNQLETSNVDMRSREMYN 472
Db 331 -----WSATOSSGVALLTGATGNGFTLNGALEASNVLSKELVN 371
QY 473 MIIQRFQMNKSVTTADTLMQKALELK 501
Db 372 MIVAQRNYQSAQRIKIQDQILNTLVNLR 400

RESULT 9

US-10-009-823A-9
; Sequence 9, Application US/10009823A
; Publication No. US20030157120A1
; GENERAL INFORMATION:
; APPLICANT: Panaccio, Michael
; APPLICANT: Rosey, Everett Lee
; APPLICANT: Sinistaj, Meri
; APPLICANT: Hasse, Detlef
; APPLICANT: Parsons, Jim
; APPLICANT: Ankenbauer, Robert G.
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE
; TITLE OF INVENTION: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES
; FILE REFERENCE: DAV150.001APC
; CURRENT APPLICATION NUMBER: US/10/009,823A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/AU00/00437
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/133,973
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Salmonella typhimurium
US-10-009-823A-9

Query Match 12.9%; Score 333; DB 14; Length 401;
Best Local Similarity 23.1%; Pred. No. 1.5e-18;
Matches 124; Conservative 64; Mismatches 147; Indels 202; Gaps 17;
QY 11 GAKTHSTGLTVSNNIANANTIGKQQVVFQDLFSODLAIGTSGOPNQAGMGAQVGS 70
Db 21 GKVAGTNLDVIGNNIANSATYFSGTI-----TQFTDGTITNG-----TGLDVAIA- 70
QY 71 VRTFTQGAFFGNSVTDLAIGGKFFQ-VLEDKVHYTAGNFRFTQDGFNDPSGFTL 129
Db 71 -----SFADMPASQNGFFELVDNSGVFSYRNGQFKLDENRNLVNMQG--- 113
QY 130 MGSRISSNPNTKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTDKTO---SEAN 186
Db 114 -----QINLNSDTP--VMQLTGTPA*GTPTTQOGANPSKTFPSVSDAD 155
QY 187 PYFALLESKNGGTPTPI-----STGNYSYAQPMRYVDQGNSHDITVYFDGAPSS 236
Db 156 SY-----NKPATIPNTLMAAKSTTTASKMGFTVTVDSQNAHDMKVYP-----VK 202
QY 237 TGSKTFFVLVAMNPSDESAAGTDSAGLLMSGTMTSSNGELKNMTAFTPTGSATKDLN 296

Db 203 TKDNEWAVYTGTVNHDSSDPAATAPITTTASTTLKFNENGILES-----TGT----- 250
QY 297 ANQAPLVNGLPOFSANFVGAGIQPLTDLFGIKSQONWAGAPASAAAIGTDIGKLPMM 356
Db 251 -----ING----- 253
QY 357 PIQTSSGNSTANGSSSTRYSQDGYPGQLVDVTTITSEGKLOGKYSNSQVDFYN----- 412
Db 254 -----ATATFSL-----NNDGTVVGNYSNEQ--EFLNSMQQ 284
QY 413 -----IPLARFTSEDLRREGNNH-----YSATLDSGGEFFG 444
Db 285 NTGQVLGQIVLANFANNEGLASQGDNIIVATNONGYKPGDLVSVQIVWAATQASGVALLG 344
QY 445 LPGTSNYKGLSVNQLETSNVDMRSREMYNIIQRFQMNKSVTTADTLMQKALELK 501
Db 345 TAGSGNFGLTNGALEASNVLSKELVNIIVAQRNYQSAQRIKIQDQILNTLVNLR 401

RESULT 10

US-10-335-977-4818
; Sequence 4818, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4818:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...221
; SEQUENCE DESCRIPTION: SEQ ID NO: 4818:
US-10-335-977-4818

Query Match 11.4%; Score 296; DB 15; Length 221;
Best Local Similarity 33.2%; Pred. No. 6.6e-16;
Matches 78; Conservative 37; Mismatches 98; Indels 22; Gaps 3;

QY 272 TFSNGELKNTAFTPGSATKDLNAWQAPLVNGLPQPSANFVGAGIPLTLDFGKISQ 331
DB 5 SYRSEGAEMFRVIVPEPGLVGSARPNVFGGRILHFNNDGSLAGMPPPLQFPDKX- 63
QY 332 QNMWAGAPASAAAIGTDIGKLPNMPTQSGNS---TARNSSSTRYRYSQDGPQGD 387
DB 64 ----GADA-----PQRINLAPGSSGDFDLTSVDKISYEALEQNGYQAGDL 106
QY 388 VDTIISSEGLKQKYSNSQVVDYNTPLAFTSEDLRREGNNHYSATLDSGPFEGLP 447
DB 107 MDVRFSDGVLGAFNSGRTLAALANFANDAGLQALGGVNFQVSGTNGSQAALGAAN 166
QY 448 TSNYKGLSVNQLTSNDMSRMVNMIIIRGFQWMSKSVTTADTMLOKALELKR 502
DB 167 TGRGSISSGKLESSNVLSRLNLIVQRGFQANSKAVTTSDQILNTLLNLQ 221

RESULT 11
US-10-335-977-4763
; Sequence 4763, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4763:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 262 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...262
; SEQUENCE DESCRIPTION: SEQ ID NO: 4763:
11.1; Score 288; DB 15; Length 262;
Best local similarity 21.0%; Pred. No. 3.8e-15;
Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;

QY 1 MWGSLFICATGKMTHTGLGTVSNANANTIGYKQQVVFQDLFSQDLAIGTSGQGN 60

DB 1 MLRSIYATSCMLAQOHTDITTSNNIANVNTGFKSRADFNDFYQAMQVAGINTSNTT 60
QY 61 -----QACMAQVGSVRTIIFTOGAFBPCNSVTDLAICGKGFQVYTLDE-KVHYTRAGNF 113
DB 61 LSPDGMVGLGVRPSAITKMFSGSPKRETNLDIAITGKGFQVQQLPDGTTATRSNGF 120
QY 114 RFTQGFNDSPGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVYN 173
DB 121 KLDEQNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 150
QY 174 LDSDTKTQSEANPYFALLEKSWKNGTTPISTNSYSAQPMRVVDDQGNSHDITVYFDGA 233
DB 151 IG-----VDGT 156
QY 234 PSSTGSKTFEYLAMNPSEDGSAASGSDSAGLLMSGTMFTSSNGELKNMTAFTPGSATK 293
DB 157 VSVT----- 160
QY 294 DLNAWQAPLVNGLPQPSANFVGAGIPLTLDFGKISQNMWAGAPASAAAIGTDIGKLP 353
DB 161 ----- 160
QY 354 SMIPTQSSGNSTARNGSSSTRYSQDGPQGDLDVDTITSEGLKQKYSNSQVVDYNI 413
DB 161 -----QGLQTTNSVIG--QI 173
QY 414 PLARFTSEDLRREGNNHYSATLDSGPFEGLPQPSNYGKLSVNQLTSNDMSRMVNM 473
DB 174 TLANFVNPAHLHMGDNLFSTINASGDAIVGNPDSQGLGKLRQGLFELSIVRLVEEMTDL 233
QY 474 IIRGFQWMSKSVTTADTMLOKALELKR 502
DB 234 ITAQRAYEANSKSIQTADAMLQTVNSLKR 262

RESULT 12
US-10-335-977-4764
; Sequence 4764, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4764:
; SEQUENCE CHARACTERISTICS:

LENGTH: 271 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...271
SEQUENCE DESCRIPTION: SEQ ID NO: 4764:
US-10-335-977-4764

Query Match 11.1%; Score 288; DB 15; Length 271;
Best Local Similarity 21.0%; Pred. No. 4e-15; Indels 254; Gaps 9;
Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;
QY 1 MMGSLFIGATGKHTSTGLTVSNANANTIGYKQQQVFDLFSQDLAIGSTGSGPN 60
DB 10 MLRSLSYATSGMLAQOHTDITSSNNIANVNTGPKSRADFNDFYQAMQVAGINTSNTT 69
QY 61 -----QAGMAQVGSVRTITQGAPEPNSVTDLAIGCKGFFQVLTLED-KVHYTRAGNF 113
DB 70 LSPDGMVEGLGVRPSAITKMFSGQSPKETTENNLDIAITGKGFQVQLPDGTTAYTRSGNF 129
QY 114 RFTQDGLNDPFGFTLMGSRISNNPNKKEITLPIQLDFNDPTVAKSPAKTSTALNAVN 173
DB 130 KLDEQGNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 159
QY 174 LGDSTDKTQSEANPYFALLESWKNGTPTSTSYVAQPMRVYDQGNSHDITVYFDGA 233
DB 160 IG-----VDGT 165
QY 234 PSSTGSKTFEYLIVAMNPSEDGSAAGTDSAGLLSGMTWTFSSNGELKNMTAFTTGSATK 293
DB 166 VSVT----- 169
QY 294 DLNAMQAPLVNGLPQFSANFVGAGIQPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLP 353
DB 170 ----- 169
QY 354 SMMPIQTSSGNSTARNGSSSTRYSQDGYPOGDLVDVITITSEGKLGKYSNQVDFYNI 413
DB 170 -----QGLQTTSNVIG--QI 182
QY 414 PLARFETSDGLRREGNNHYSATLDSGGPEFGLPGTSGYKLSVNQLETNSVDMRSRMVNM 473
DB 183 TLANFVNPAFLHSMGDNLFSTINASGDAIVGNPDQGLKRGFLELSNVLVEEMTDL 242
QY 474 IIQRGFQWNSKSVTTADTLMQKALELKR 502
DB 243 ITAQRAYEANSKSIQTADAMLQTVNSLKR 271

RESULT 13
US-10-335-977-4765
; Sequence 4765, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4765:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...279
SEQUENCE DESCRIPTION: SEQ ID NO: 4765:
US-10-335-977-4765

Query Match 11.1%; Score 288; DB 15; Length 279;
Best Local Similarity 21.0%; Pred. No. 4.2e-15;
Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;
QY 1 MMGSLFIGATGKHTSTGLTVSNANANTIGYKQQQVFDLFSQDLAIGSTGSGPN 60
DB 18 MLRSLSYATSGMLAQOHTDITSSNNIANVNTGPKSRADFNDFYQAMQVAGINTSNTT 77
QY 61 -----QAGMAQVGSVRTITQGAPEPNSVTDLAIGCKGFFQVLTLED-KVHYTRAGNF 113
DB 78 LSPDGMVEGLGVRPSAITKMFSGQSPKETTENNLDIAITGKGFQVQLPDGTTAYTRSGNF 137
QY 114 RFTQDGLNDPFGFTLMGSRISNNPNKKEITLPIQLDFNDPTVAKSPAKTSTALNAVN 173
DB 138 KLDEQGNLVTSEGYLLI-----PQI---TL-----PEDTTQ-----VN 167
QY 174 LGDSTDKTQSEANPYFALLESWKNGTPTSTSYVAQPMRVYDQGNSHDITVYFDGA 233
DB 168 IG-----VDGT 173
QY 234 PSSTGSKTFEYLIVAMNPSEDGSAAGTDSAGLLSGMTWTFSSNGELKNMTAFTTGSATK 293
DB 174 VSVT----- 177
QY 294 DLNAMQAPLVNGLPQFSANFVGAGIQPLTLDFGIKSOQNMWAGAPASAAAIGTDIGKLP 353
DB 178 ----- 177
QY 354 SMMPIQTSSGNSTARNGSSSTRYSQDGYPOGDLVDVITITSEGKLGKYSNQVDFYNI 413
DB 178 -----QGLQTTSNVIG--QI 190
QY 414 PLARFETSDGLRREGNNHYSATLDSGGPEFGLPGTSGYKLSVNQLETNSVDMRSRMVNM 473
DB 191 TLANFVNPAFLHSMGDNLFSTINASGDAIVGNPDQGLKRGFLELSNVLVEEMTDL 250
QY 474 IIQRGFQWNSKSVTTADTLMQKALELKR 502
DB 251 ITAQRAYEANSKSIQTADAMLQTVNSLKR 279

RESULT 14

US-10-009-823A-11		US-10-009-823A-12	
; Sequence 11, Application US/10009823A		; Sequence 12, Application US/10009823A	
; Publication No. US20030157120A1		; Publication No. US20030157120A1	
; GENERAL INFORMATION:		; GENERAL INFORMATION:	
; APPLICANT: Panaccio, Michael		; APPLICANT: Panaccio, Michael	
; APPLICANT: Rosey, Everett Lee		; APPLICANT: Rosey, Everett Lee	
; APPLICANT: Sinistaj, Meri		; APPLICANT: Sinistaj, Meri	
; APPLICANT: Hasse, Detlef		; APPLICANT: Hasse, Detlef	
; APPLICANT: Parsons, Jim		; APPLICANT: Parsons, Jim	
; APPLICANT: Ankenbauer, Robert G.		; APPLICANT: Ankenbauer, Robert G.	
; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE		; TITLE OF INVENTION: LAWSONIA DERIVED GENE AND RELATED FLGE	
; FILE REFERENCE: DAVI150.001APC		; FILE REFERENCE: DAVI150.001APC	
; CURRENT APPLICATION NUMBER: US/10/009, 823A		; CURRENT APPLICATION NUMBER: US/10/009, 823A	
; CURRENT FILING DATE: 2002-08-13		; CURRENT FILING DATE: 2002-08-13	
; PRIOR APPLICATION NUMBER: PCT/AU00/00437		; PRIOR APPLICATION NUMBER: PCT/AU00/00437	
; PRIOR FILING DATE: 2000-05-11		; PRIOR FILING DATE: 2000-05-11	
; PRIOR APPLICATION NUMBER: US 60/133,973		; PRIOR APPLICATION NUMBER: US 60/133,973	
; PRIOR FILING DATE: 1999-05-13		; PRIOR FILING DATE: 1999-05-13	
; NUMBER OF SEQ ID NOS: 13		; NUMBER OF SEQ ID NOS: 13	
; SOFTWARE: FastSeq for Windows Version 4.0		; SOFTWARE: FastSeq for Windows Version 4.0	
; SEQ ID NO 11		; SEQ ID NO 12	
; LENGTH: 480		; LENGTH: 360	
; TYPE: PRT		; TYPE: PRT	
; ORGANISM: Aquifex aeolicus		; ORGANISM: Vibrio parahaemoliticus	
US-10-009-823A-11		US-10-009-823A-12	
Query Match 10.9%; Score 282; DB 14; Length 480;		Query Match 9.0%; Score 233; DB 14; Length 360;	
Best Local Similarity 22.9%; Pred. No. 2.9e-14;		Best Local Similarity 22.0%; Pred. No. 1.9e-10;	
Matches 135; Conservative 73; Mismatches 160; Indels 222; Gaps 28;		Matches 110; Conservative 59; Mismatches 127; Indels 204; Gaps 21;	
Qy 1	MMGSLFTGATGM-----KTHSTGLGTV-----SNNIANANTIGVKQOQVFPQDL 44	Qy 11	GMKTHSTGLGTVSNNIANANTIGVKQOQVFPQDLFSQDLAIGSTGSGQPNQAGMAQVGS 70
Db 1	MLRSFYNAITGMDVSRGTVKTITFGAGAVVDSFALDVTSDNLNANVTGFKKSR----- 54	Db 20	GUEVASTELNITSHIANASTYGFK-----GARISSQ 50
Qy 45	FSQDLAIGSTGSGQPNQAGMAQVGSVRTITFTQGAPEPGNSVTDLAI----- 91	Qy 71	VRTITFTQGAPEPGNSVT-----DLAI-----GGKGFQVTLK-----DKVHYTRAGNFR 114
Db 55	-----TQKVWITGSEFKQTEITTDLAIPFDQMVSVQVWG 88	Db 51	-----NFKNGSITGTGRMDLAITEFAAVYNNNGSGFF-VTKDHMGQTLYTRSGVFG 101
Qy 92	-----GKGF--QVTLKDVHYTRAGNFRPTQDGLNDPSGFTLMGSRISNNPNKKET 144	Qy 115	FTQDGLNDPSGFT-----LMGSRISNNPNKKETLEPIQLDEN--DPTVAKSPA 162
Db 89	LNTTTEGKALFILRDVLNQTYYTRDRGRFRINREGYLINFG----- 130	Db 102	TKSKNFVYFANNNGVANFNDASAKAALQGVSDSNNNLMTGSDIKAVTPPDADP----- 154
Qy 145	LEPIQLDNPTVAKSPAKTALNAVNLGSDSTKQSEANPYFALLEBSWKN----- 198	Qy 163	KTSTALNAVNLGSDSTKQSEANPYFALLEBSWKNGGTPPITSTSYSAQPMRVYDQGN 222
Db 131	--PTNLDERAPILYVQGFKNVPTGEVT--GTQIDQTTTFNP-----LDSFTNYLEDIR 182	Db 155	--TSFNSVGNIQVSTSSLNKA-----TDKLDFTSYTTQVYDVLGN 192
Qy 199	---GTPPISTNSYSAOP---MRVYDQGNSHDITVYFDGAPSGTG----- 238	Qy 223	SHDITVYF-----DGA-----PSST-----GSKTPEYLVAMNP---SEDSASAGTDSAGL 265
Db 183	VETQIPPKATGEI-YFNPRYTLTIYDSLREVPADIYF-----VKTGNTQWKVYFTLPTFA 237	Db 193	SHTVTYFTKTADNAWEVNVPTGSEFNQVDGKGT--PVVSTIPEVTNKDGTLLAAFPAPAG- 249
Qy 239	SKTPEYLVAMNPSEDGSA-----SGTDSAGLLMSGTMTSSNGELKNMTAFTPTGS 290	Qy 266	LMSGTMTSSNGELKNMTAFTPTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQPLTLD 325
Db 238	SKTLF-----DPSTLASKERYINVDWNGDDDKGLV-----YIPGGD 276	Db 250	--ANAMSVVDINEDNGMVYATYNGOSLK-----GSTQFGAQLQG--QVVLAD 292
Qy 291	ATKDLNAWQAPLVNGLPQFSAN--FVYAGIQPLTLDGFKSKQNMWAGAPASAAAGT- 347	Qy 326	FGKSKQNMWAGAPASAAAGTIDGKLPSPMMPITQSSNGSTARNSSSTRYSQDGYQGG 385
Db 277	IVQD-----TANQKFLYTDIVFLDL--FNDQVHI-----ADNGTF 309	Db 293	FA-----NTQGLAKVSGFGVSTN-----SPNGYTSG 318
Qy 348	DICKLSPMMPITQSSNST-----ARNGSSRFR-RYSQDG-----YPOGDLVDV-- 390	Qy 386	DLAVDTYITSEGLKQCKYNSQVWDFYNIPLARFTSDDLREGNNHYSATLDSGGPEPGL 445
Db 310	SEVLDLTPESGSEINDPNDTESYLNKLGAKLSETKIKLSEGVVGVVYNSGETIYVGE 369	Db 319	ELAGVRVA-----WTQSSSGAPIMGV 340
Qy 391	TITSEGLKQCKYNSQVWDFYNIPLARFTSDDLREGNNHYSATLDSGG----- 440	Qy 446	PGTSNYGKLSVNQLETSNVD 465
Db 370	GILQNNVQNSYITQHALPYRLAALQFTDPEELVKKGSDFVVTNDQGVARGELIDL 428	Db 341	PGSGTGLNLTGALBGSNVD 360
Qy 441	-----PEPGLGTSNYGKLSVNQLETSNVDMSRENMWIIQRCFQM 482		
Db 429	VLYASVKTPTILLPGGSGN--KIRSAVVMGNGVNDIAKEFINLITACQTYQV 476		
RESULT 15		RESULT 16	
US-10-335-977-4840		US-10-335-977-4840	
; Sequence 4840, Application US/10335977		; Sequence 4840, Application US/10335977	
; Publication No. US20040052799A1		; Publication No. US20040052799A1	
; GENERAL INFORMATION:		; GENERAL INFORMATION:	

APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 4840:

SEQUENCE CHARACTERISTICS:

LENGTH: 152 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...152

SEQUENCE DESCRIPTION: SEQ ID NO: 4840:

US-10-335-977-4840

Query Match 7.8%; Score 201.5; DB 15; Length 152;
Best Local Similarity 34.6%; Pred. No. 2e-08;
Matches 47; Conservative 26; Mismatches 52; Indels 11; Gaps 3;

QY 378 SQDGYQCDLVDTTTSCKLQGYNSQVVDYFNIPLAFTSDGLRREGNNHY---SA 434

Db 17 SQDGPGRIFEDMRTEENGVSIFAFNGVPEPVARIGLAFTNDQGLRKIGGNLYEMQEG 76

QY 435 TLD-----SCGPEPL--PQTSYKGLSVNQLTSDVMSREVMNMIITQGFQMSKS 486

Db 77 TINGENRPLSGNPILGWDDEKLFKGRKIRKYLETSNVNAGNALTNLILMQGYSMNARA 136

QY 487 VTTADTLMQALELKR 502

Db 137 FGAGDDMIKEAISLKK 152

RESULT 17

US-10-335-977-4844

Sequence 4844, Application US/10335977

Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

RELATING TO HELICOBACTER PYLORI FOR

DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 4844:

SEQUENCE CHARACTERISTICS:

LENGTH: 464 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...464

SEQUENCE DESCRIPTION: SEQ ID NO: 4844:

US-10-335-977-4844

Query Match 7.3%; Score 189.5; DB 15; Length 464;
Best Local Similarity 28.9%; Pred. No. 9.9e-07;
Matches 59; Conservative 29; Mismatches 81; Indels 35; Gaps 6;

QY 1 MMGSLFIGATGKTHSTGLGVSNNIANANTIGYKQQVWFQDLFSQDLAIGSTGSGQPN 60

Db 1 MNDTLNAYSGIKTHQFGIDSLSNNIANVNTLGYRSNDPEKTLFSSHLDAIKNVAVAN 60

QY 61 QAGCAQVGSVVTITQ-GAPEGNSVTDLAIGGKGFQV-----TLEDK 104

Db 61 DRNYGV-TSGNVLSNKDGYNPSEGEFHMAYQKGKGVFNKNGENTINKDGFSGKQD 119

QY 105 VHYTRAGNFRFTQDGLNDPSGFTLMG-----SRISNNPNIKK---ETLEPIQL 150

Db 120 NFLTRAGNFARDADGLVTPGYYVYGIDLKKIKDGLNSTARDEDIEKLHGNTLSPLQI 179

QY 151 DPNDPTVAKSPAKTSTALNAVNL 174

Db 180 ----PQDLTYQPVLSKTVNISVNL 199

RESULT 18

US-10-009-823A-13

Sequence 13, Application US/10009823A

Publication No. US20030157120A1

GENERAL INFORMATION:

APPLICANT: Paraccio, Michael

APPLICANT: Rosey, Everett Lee

APPLICANT: Sinistaj, Meri

APPLICANT: Hasse, Detlef

APPLICANT: Parsens, Jim

APPLICANT: Ankenbauer, Robert G.

TITLE OF INVENTION: LAMSONIA DERIVED GENE AND RELATED FLGE

TITLE OF INVENTION: POLYPEPTIDES, PEPTIDES AND PROTEINS AND THEIR USES

FILE REFERENCE: DAVI150.001APC
CURRENT APPLICATION NUMBER: US/10/009,823A
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: PCT/AU00/00437
PRIOR FILING DATE: 2000-05-11
PRIOR APPLICATION NUMBER: US 60/133,973
PRIOR FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 13
LENGTH: 370
TYPE: PRP
ORGANISM: Borrelia burgdorferi
US-10-009-823A-13

Query Match 7.2%; Score 186; DB 14; Length 370;
Best Local Similarity 21.9%; Pred. No. 1.4e-06;
Matches 98; Conservative 57; Mismatches 143; Indels 150; Gaps 19;

1 MMGLFTGATGMKTH-----STGLGTVSNINANANTIGYKQOQVWFQDLF 45
1 MMSLYSGVGLQHOGTNGKQVGLGNVASTMDVVGNNIANVNTIGFKK----- 51
46 SQDLAIGSTGSGQGNAGMAQVSGVVRITFTQGAPEPNSVTDLAI----- 91
52 -----GRDITHTQGAFOSTOKASDLGVVNFQDMISQISGA 88
92 -----GKGFQVLTEDKHYTRAGNFRFTQDGLNDP-SGFTLMGSRISNNENIKK 142
89 SRPTDARGSGNGFFILKEGPNLFTRAGAFDVSDRHLVNPANGACNLDKRLVRIQGWAA 148
143 ETLLEPIQL--DFNDPTVAK--SPAKTSTAL-NAVNLGSDTDKTOSEANPYFALLBSWK 197
149 RDLGKRVINTASDPLIOEGANPADIARGIEDLIIPIDGKXGAKSTKVTFF-----TWVV 203
198 NGTPISTSYSAQPMRVYD-QQGNSHDITVYFDGAPSGTSGKTFEYLVAMNPSEDGSA 256
204 NKSLYDSFGVSVLELRVWLDLNTPLNNAIV-LNGQP-----LNGEONG-- 251
257 ASGTDSAGLMSG-TMTFSSNGELKNNMTAFTPTGSATKOLNAPAPLVNGLPQFSANFV 315
252 ----DILQIPISNFTLGFDEGALAS-----TFNVL 278
316 GAGIQPL--TLDFGKSCQNNWAGAPASAAAIGT-----DIGK--LFSMMP 357
279 GANVTEGETQTVNLKDDQGVIVGIYSGIRLGTGVSYTDSITQFADRDILGKIALASFM- 337
358 IQTSSGNSTARNSSSTRYSQDGPQ 385
338 ----NPGDLQNALISSTKAIIDQGYGMG 361

RESULT 19
US-10-369-493-4713
Sequence 4713, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 4713
LENGTH: 655
TYPE: PRP

ORGANISM: Burkholderia fungorum
US-10-369-493-4713
Query Match 7.0%; Score 180.5; DB 14; Length 655;
Best Local Similarity 22.4%; Pred. No. 8.9e-06;
Matches 138; Conservative 75; Mismatches 225; Indels 178; Gaps 30;

1 MMGLFTGATGMKTHSTGLGTVSN--IANANTIGYKQOQVVFQ-----DLFQD 48
98 LIGSPTAGIASAITSYFICLQNVSNASSLATRQATMGSAQTLVQINAAQQYDALRQS 157
49 L--AIGSTGSGQPNQAGMAQV-GSVRTIFFTQGAPEPNSV---TDLAIGKGFQVTLT 102
158 VNTQLTNTVQSINSYSQIAQLNGGIAQASTQG--QPPNQLLDQDLAVSN----- 206
103 DXVHYTRAGNFRFTQDGLNDPSPGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKPSA 162
207 -----LSQLIGVNVVNS--NGSYVFMGSGPLVSTSYNIGTAPS 246
163 KTSTALNAVNLGSDTDKTOSEANP-----YFALLBSWKNGTPTPI 203
247 TGDTSLSVOYLG-----QAGANPAATFQNLDPDSKIDGTLGGLLAFRSQTLDPGEAQL 300
204 STSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTSGKTFEYLVAMNPSEDGSAASGTDSA 263
301 GAIAVSFS--AQVNAQNGLGITLAGAKGALFSVGGPT---VYANTQNTGNASLVNSFA 354
264 GLLM--SGTMTFSSNGELKNNMT--AFTPTGSATKOLNAPAPLVNGLPQFSANFVGAG 318
355 DATQPTTGDYILAYNGTYYTLTDNSTGTIVGSAT---NLSQP-----NFSITG 403
319 IQPLDLDFGKSCQNNWAGAPASAAAIGTIDKGLPSMMPI-----QTSNGSTARNSS-- 371
404 ----TNAGDSFTVFTREGALNSFATATTDASIAAAAAPVLAASANTGTATITQGTVT 459
372 -----SSTRYSQD-----GYPOQDLVDV-----TITSE 395
460 AGYTPNPTTLLSYDGTGLSGFPAGSTTVVAGTPTTITASAATVVPYSATGATLFIN 519
396 GKLOGKYSQNVDFYNIPLA--RFT--SEDLRREGNN-----HYSATLDSGGPEFGLP 446
520 NATAGQNNVSVT-ISNAPAGDKFTIGPNTGATNDGRNALALSNLSTAKAMSGGTVILT 578
447 G-----TSNYGKLSVNOLETNS-----VDMSEVMNMIIOGFQM 482
579 GAYANYVNNIG-OTNQITQTSSTAOSSLVQITTAQQSVSGVNVNINEEANLLQYQQLYQA 637
483 NSKSVTTADTMLQKAL 498
638 NSKVIQTAQTLFQTIL 653

RESULT 20
US-10-369-493-7473
Sequence 7473, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 7473
LENGTH: 655
TYPE: PRP

Db 1011 IKAGVDGENSDSDATNANLTIKTELK-----TQDLNI-----SGFNKAEITAKDGS 1059
Qy 71 VRTIFTOGAPEPGNS-----VTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLND 123
Db 1060 DLTIGNTNSADGNAKVFQVQVDSKISADG-HKVTLHSKVE--TSGSNNTEDSSDNN 1116
Qy 124 PSQFTLMGSRISNNPIKKELEPIQLDFNDPFAKSPAKTSTALNAVNLGSDTDKTS 183
Db 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGBITTKTGTINATGNVEITATQGS 1170
Qy 184 EANPYFALLESWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVFDGAPSGTG 238
Db 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN--SCALTTLA 1214
Qy 239 SKTFEYLVANMPSEDSGAASGTDAGLL--MSGTMTFSSNGELKNMTAFTPTGSAKDL 295
Db 1215 GSTIKGFTSVTSSQSGDIGTISGGIVKATESLTQTSNKIKATTGEANVTSATGTI 1274
Qy 296 NAMQAPLVNGLPQFSANFVAGIQPLTLDGFIKQNMWAGAPASAAIGTDIGKLPSP 355
Db 1275 GGTISGNTVN-----VTANAGDLTVGNGAE-----INATEGAATLTSSGKL---- 1316
Qy 356 MPIQTSSGNTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL--OQK----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAAVNTLTGTLTVKGSNNINATSGTL 1372
Qy 402 YSNSQVDFNIPILARTSEDLRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNQLET 461
Db 1373 VINAKDAELNGAALGNHTVVNATNANGSGSVIATSS--RVNITG---DLITINGL-- 1423
Qy 462 SNVDMREVMNMIIOGFQFQNSKSVTTADTLMQKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEIVEAKR 1461

RESULT 23
US-10-193-764-63
; Sequence 63, Application US/10193764
; Publication No. US2003013943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 1536.
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-63

Query Match 6.7%; Score 173.5; DB 14; Length 1536;
Best Local Similarity 22.8%; Pred. No. 0.00012; Indels 100; Gaps 24;
Matches 119; Conservative 80; Mismatches 222;
Qy 12 MKTSTGLTGVSNNIANAN-TIGYKQVQVFDLFSQDLAIGSTGSPQAGNAGVGS 70
Db 1011 IKAGVDGENSDSDATNANLTIKTELK-----TQDLNI-----SGFNKAEITAKDGS 1059
Qy 71 VRTIFTOGAPEPGNS-----VTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLND 123
Db 1060 DLTIGNTNSADGNAKVFQVQVDSKISADG-HKVTLHSKVE--TSGSNNTEDSSDNN 1116
Qy 124 PSQFTLMGSRISNNPIKKELEPIQLDFNDPFAKSPAKTSTALNAVNLGSDTDKTS 183

Db 1117 -AGLTIDAKNVTNNITSHKAVSI-----SATSGBITTKTGTINATGNVEITATQGS 1170
Qy 184 EANPYFALLESWKNGT-----PPISTSNYSYAQPMRVYDQGNSHDITVFDGAPSGTG 238
Db 1171 -----ILGGIESSGSVTLTATEGALAVSNIS-----GNTVTVTAN--SCALTTLA 1214
Qy 239 SKTFEYLVANMPSEDSGAASGTDAGLL--MSGTMTFSSNGELKNMTAFTPTGSAKDL 295
Db 1215 GSTIKGFTSVTSSQSGDIGTISGGIVKATESLTQTSNKIKATTGEANVTSATGTI 1274
Qy 296 NAMQAPLVNGLPQFSANFVAGIQPLTLDGFIKQNMWAGAPASAAIGTDIGKLPSP 355
Db 1275 GGTISGNTVN-----VTANAGDLTVGNGAE-----INATEGAATLTSSGKL---- 1316
Qy 356 MPIQTSSGNTARNGSSSTRYSQDGYPOGDL--VDVTITSEGL--OQK----- 401
Db 1317 -----TTEASSHITSAGQVNLQAQDSVAGSINAAVNTLTGTLTVKGSNNINATSGTL 1372
Qy 402 YSNSQVDFNIPILARTSEDLRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNQLET 461
Db 1373 VINAKDAELNGAALGNHTVVNATNANGSGSVIATSS--RVNITG---DLITINGL-- 1423
Qy 462 SNVDMREVMNMIIOGFQFQNSKSVTTADTLMQKALELKR 502
Db 1424 -NI-ISKNGINTVLL-KGVKIDVKYIQPGIASVDEIVEAKR 1461

RESULT 24
US-10-282-122A-77944
; Sequence 77944, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77944
; LENGTH: 3705

```
; TYPE: PRT
; ORGANISM: Versinia pectis
US-10-282-122A-77944

Query Match
Best Local Similarity 6.7%; Score 172.5; DB 15; Length 3705;
Matches 119; Conservative 58; Mismatches 180; Indels 191; Gaps 24;

QY 3 GSLFTGATGKTHSTGLGTGVSNNIANANTIGYKQOVVFDLFSQDLAIGTSGSQGPNQA 62
Db 2359 GELIILGANGAGCQTSLLNIAG--ASANINGRQ-----TVGAVTNGVTLL 2403
QY 63 GNGAAGVSVRIFTQGAPEGPNVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLN 122
Db 2404 GNGBELTSTDILNTGMINVTGILNLENGGASSISGGL-----TGNGILN 2449
QY 123 DPSG--FTLMGSRISNNPNIKETLEPIQLDFNDFTVAKSPAKTALNAVNLGSDTKT 181
Db 2450 IKGSDFTI---SIDNN-----GLAQTNISDGSV 2477
QY 182 QSEANPYEALLESKMGNGTTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGAPSGTSGKT 241
Db 2478 -----LNGGTIIIGCN-----LGSSV 2494
QY 242 FEYLVAMPSEDGSAAS-----GT--DSAGLLMSGTWTFS-----SNGEL-----KNMT 283
Db 2495 IDVLGDLNLVADNSLANVISGDGTINTTATVTLSGNSFSFGAHQIGTNGELTVGQASNLG 2554
QY 284 AFTPTGSAKDNLNAQWAPLNGLPQFSANFVGAGIQPLTLD----- 325
Db 2555 A-----SSATVNLGLTSLHILNGVSEIAN--VLSGVAGSVIILGGADTALTANNISFLG 2609
QY 326 -----FGIKSQONWAGAPASAAAIGTDIGKLPMMPIOTSSGNSTARNSSSTR 376
Db 2610 QVALAGNSKLTAVASTNLL--GASSVALAGT--GDTLSLGFNGTFCNSVTGSGVLQVT-- 2664
QY 377 YSQDGYPGQDLVDVTIISGKLGKYSQVDFYNIPL-----ART-----SEDGLRREG 428
Db 2665 -----DDEVTLSNGV-----SNAVTDIDATLNLDDIALLFNHALTGNNLVA 2711
QY 429 NNHYSATLDSGPFEG--LPGTSNYKLSVNOLETNSVDMGSRVMNIIIQGFQMSKSV 487
Db 2712 KNDASTAFDFGATVGGAFTGVN-----LNNSTFDLSGNNTT--VLAQATLKLSSGNL 2762
QY 488 TTADTMLQ 495
Db 2763 TSVGNVQ 2770

RESULT 25
US-10-335-977-4817
; Sequence 4817, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977

; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4817:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...91
; SEQUENCE DESCRIPTION: SEQ ID NO: 4817:
US-10-335-977-4817

Query Match
Best Local Similarity 6.4%; Score 166; DB 15; Length 91;
Matches 33; Conservative 20; Mismatches 33; Indels 0; Gaps 0;

QY 400 GKYSNSQVDFYNIPLAFSTEDGLRREGNHHYSATLDSGPFEGLPFTSNYKLSVNQL 459
Db 5 GAFSGRTLLAQAVALANFANDAGLQALGDVNFQGTGNSGOALIGAANTRRGISGSKL 64
QY 460 ETSNVDMGSRVMNIIIQGFQMSK 485
Db 65 ESSVDLSRSLTNLIVVQGFKRLK 90

RESULT 26
US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4

Query Match
Best Local Similarity 6.4%; Score 166; DB 14; Length 2468;
Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;

QY 20 GTVSNNIANANTIGYKQOVVFDLFSQDLAIGTSGSQG-----PNCAGMGAQVGSV 71
Db 1298 GTVNAVA-----QDPA-GNTGPGSTTVDVAPNTEVVPNSGNL 1337
QY 72 RTITQGAPEGPNVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGLNDSGFTLMG 131
Db 1338 ----LNGTAEPGSTVTLTDGNGNPIGQTADG-----SGNWSFTPGSQLPNGTVVAVTA 1387
```

QY 132 SRISNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPYFAL 191
Db 1388 SDAAGNTSLPATT---TVDSSLPSIPQVDPSPNGSVISGTADAGNTIIITDNGNPIGOV 1443
QY 192 LESWKNG--TPPISTSNYSYAQPMRVYDQGNSHDITVYFDG-APSSGSKTFFYLAVM 248
Db 1444 TADGSGNWFTFGIPDPGTGVNVVARSNVDSPAPVITVDGVAPAP---VI 1494
QY 249 NPSEDGSAASGTDAGLLM-----SGTMFTS-----SNGELKNMTAFT 286
Db 1495 DPS-NGTEISGTAEAGATVILTDGNGNPIGOATADGSGNWTTPSTPLANGTVINAVAQD 1553
QY 287 PTG-----SATKDLNMQAPLNV-----GLPQFSANFV-----GAGIQPLTLDG 327
Db 1554 PAGNTSGPASVTVDALIA-PPAPVINPSNGVVISGTAEAGATVILTDGNGNPIGOV 1610
QY 328 IKSQONMWAGAPASAAAGTIDGKL-----PSMMPITQSSGNSTA 367
Db 1611 ---GSGNWFTFGTPLANGSVINALAQDAAGNNSPTSATVDSLAPAPVIDPNSGVIA 1667
QY 368 RNGSSSTRYSQD--GYQGDLDVDTITSEKLGKQKYSNQVDFYNIPLARFTSEDGLR 425
Db 1668 GTAEAGATVILTDGNGNPIGO-----VTADG--SGNWSFTPGTPLSLNGTVINAVAQDA-- 1718
QY 426 REGNNH--YSATLDSGGPGLPGTSNYGKLS 455
Db 1719 -AGNTSGPVSTTVDVAVAPATFVIDPNSGVELS 1749

RESULT 27

US-10-282-122A-66335
; Sequence 66335, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 66335
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66335

Query Match 6.4%; Score 166; DB 15; Length 2468;
Best Local Similarity 21.9%; Pred. No. 0.00095;
Matches 112; Conservative 67; Mismatches 197; Indels 136; Gaps 23;

QY 20 GTVSNNIANANTICYKQQVVFQDLFSQDLAIGSTGSGQ-----PNOAGMCAQVGSV 71
Db 1298 GTVNVAVA-----QDPA-GNTGPGSGTVDVAVAPNTPVVNPNSGNL 1337
QY 72 RTITQGAFFPGNSVTDLAIGKGFVFTLEDKVHYTRAGNFRFTQDGLNDPDSGFTLMG 131
Db 1338 ---LNGTAEPGSGTIVTLDGNGNPIGOATADG-----SGNWSFTPGSGLPNGTVINAVTA 1387
QY 132 SRISNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTQSEANPYFAL 191
Db 1388 SDAAGNTSLPATT---TVDSSLPSIPQVDPSPNGSVISGTADAGNTIIITDNGNPIGOV 1443
QY 192 LESWKNG--TPPISTSNYSYAQPMRVYDQGNSHDITVYFDG-APSSGSKTFFYLAVM 248
Db 1444 TADGSGNWFTFGIPDPGTGVNVVARSNVDSPAPVITVDGVAPAP-----VI 1494
QY 249 NPSEDGSAASGTDAGLLM-----SGTMFTS-----SNGELKNMTAFT 286
Db 1495 DPS-NGTEISGTAEAGATVILTDGNGNPIGOATADGSGNWTTPSTPLANGTVINAVAQD 1553
QY 287 PTG-----SATKDLNMQAPLNV-----GLPQFSANFV-----GAGIQPLTLDG 327
Db 1554 PAGNTSGPASVTVDALIA-PPAPVINPSNGVVISGTAEAGATVILTDGNGNPIGOV 1610
QY 328 IKSQONMWAGAPASAAAGTIDGKL-----PSMMPITQSSGNSTA 367
Db 1611 ---GSGNWFTFGTPLANGSVINALAQDAAGNNSPTSATVDSLAPAPVIDPNSGVIA 1667
QY 368 RNGSSSTRYSQD--GYQGDLDVDTITSEKLGKQKYSNQVDFYNIPLARFTSEDGLR 425
Db 1668 GTAEAGATVILTDGNGNPIGO-----VTADG--SGNWSFTPGTPLSLNGTVINAVAQDA-- 1718
QY 426 REGNNH--YSATLDSGGPGLPGTSNYGKLS 455
Db 1719 -AGNTSGPVSTTVDVAVAPATFVIDPNSGVELS 1749

RESULT 28

US-10-335-977-4839
; Sequence 4839, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4839:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...124
; SEQUENCE DESCRIPTION: SEQ ID NO: 4839:
US-10-335-977-4839

Query Match 6.4%; Score 165.5; DB 15; Length 124;
Best Local Similarity 32.0%; Pred. No. 1.3e-05;
Matches 39; Conservative 24; Mismatches 48; Indels 11; Gaps 3;

QY 392 ITSEKLGKQYNSQVDFYNIPLARFTSEDLREGNNHY---SATLD-----SGGPE 442
Db 3 IEENGVISLAFSNGVVEPVAIGAXFINDGGLKIGKNLYEMOEGTINGENPLXGNPI 62
QY 443 FGL--PGTSNYGKLSVNOLETSYNDMSREVMNIIIOGFQMSKSVTTTADTLMQALEL 500
Db 63 LGWDEEGLKFGKIRHKYLETSYVAGNALTLLIMQGYSMNARAFGAGDDMIKEAISL 122
QY 501 KR 502
Db 123 KK 124

RESULT 29
US-10-335-977-4819
; Sequence 4819, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4819:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...343
; SEQUENCE DESCRIPTION: SEQ ID NO: 4819:
US-10-335-977-4819

Query Match 6.3%; Score 164; DB 15; Length 343;
Best Local Similarity 24.5%; Pred. No. 7.7e-05;
Matches 82; Conservative 52; Mismatches 129; Indels 72; Gaps 15;

QY 32 IGYKQOVVFQDLFSODLAIGS----TGSQGN--QAQMGAGVGSVRTIPTGGAPEPGNS 85
Db 3 LGFKYSRASFDVMDLSQVKLIATDPYKNGLAGQNDPSVGLGVGVDTATTKIFSQGNITDV 62
QY 86 VTDLAIGKGFQVTLDEKV--HYTRAGNFRF---TQGFNDPSGFTLM-----GSRI 134
Db 63 KTDLAIQDGFIIISPDGRTNFTRMGVSFGLAREFGYHRWACGARSVENRSDTGNGK 122
QY 135 SNNPNIKETLEPIQDNDPTVAKSPAKTST--ALNAVNLGSDTKTQSEANPYFALL 192
Db 123 SPTDALKVDNTGPLENIRIDPGMV-MPAKTSNRISMRANINAGKHADQTA--VFALD 177
QY 193 ESWK--GNGTPPISNTSYSAQPMR---VYDOQN-----SHDITVYFDGA----- 233
Db 178 SSAKTPSDGINFVDSGTNLAHVPENMGSLYNEGDALLLIENHGIWVYSKAKVXDIL 237
QY 234 PS-----STGSKTFEYLVAMNPSBDGSAAG-----TDSAGLLMSGTM 271
Db 238 PSAENSTLELNGVKISFTNDGSAVSRSTSLVAAKNAINAVKSTGTEAYLDGKQLRENTN 297
QY 272 TFSNGELKNMT-----AFTPTGSATKDLNANQ 299
Db 298 ELDGDEKLNIVTQAGTGAFANFLDGDQKVTAFK 332

RESULT 30
US-10-147-299A-4
; Sequence 4, Application US/10147299A
; Publication No. US20040058323A1
; GENERAL INFORMATION:
; APPLICANT: KO, ALBERT I.
; APPLICANT: HAAKE, DAVID A.
; APPLICANT: REIS, MITERMAYER GALVAO
; APPLICANT: MATSUNAGA, JAMES
; APPLICANT: CRODA, JULIO HENRIQUE ROSA
; APPLICANT: SIQUEIRA, ISADORA CRISTINA
; APPLICANT: RILEY, LEE W.
; APPLICANT: BAROCCHI, MICHELE
; APPLICANT: YOUNG, TRACY ANN
; TITLE OF INVENTION: PROTEINS WITH REPETITIVE BACTERIAL-IG-LIKE (BIG)
; FILE REFERENCE: 3673-19
; CURRENT APPLICATION NUMBER: US/10/147,299A
; FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1954
; TYPE: PRT
; ORGANISM: Leptospira kirschneri
US-10-147-299A-4
```


ORGANISM: Helicobacter pylori

US-09-881-752A-150

Query Match 6.3%; Score 162; DB 15; Length 1954;

Best Local Similarity 19.3%; Pred. No. 0.0014;

Matches 125; Conservative 80; Mismatches 241; Indels 200; Gaps 23;

QY 7 ICATGKTHSTGLG-TVSNNAANA-----NTIGYKQQQVVFQDLFSQDL 49

DB 410 VGSTNITAAIGVDITVSNLVNNALESQVSDSHSIARGTSTFQAIGVYSDGSSQNI 469

QY 50 -----AIGSTGQGNPQAGMAQVGSVRTIPTQGAFFPGNSVTDL 89

DB 470 SDQVAMNNSSTLIQISNLNAVPRKREIQSPSSGGLG-----TARITATLEAISSYTDI 522

QY 90 AICGCKFFQVLTEDKVHYTRAG-NPFTQDQFLNDSGFTLMGSRISNNPNKIKETL--- 145

DB 523 SVNAATLVIEVSPNTPNSVSSGLUTVFTATGVYTDGNSONLISQVTVNSNTNRATISNA 582

QY 146 -----EPIQLDFNDPT-----VAKSPAKTSTALNAVVN-----LGDSTDKT- 181

DB 583 NGTQGLAGSSVCTTNI SATLGAVTSSATLTVTNAVLNSITITPSPLSVAVGRSLNLTA 642

QY 182 -----QSEANPYFALLSWKNGTTPPISTNSYVAQPMRVYDQGNSH----- 224

DB 643 TGYSDGNSQDLATSVAMTSTDSSIVSDNASGRQQTGVAGQNTQISATLGTSSAIN 702

QY 225 -----DITYVFDGAPSTGSKTFEYLVA MNPSEDGSAASGTDSDA----- 263

DB 703 FTVSAAVLDSIQVLTEDSPAKGTST--RAIATGVFSDGSLNLSIDQVWDSSTQTNVIQL 760

QY 264 GLMSG-----TMTFSSNGELKNTATPTGSGAT-----KDLNAWQPA----- 301

DB 761 GVLETPKPKKLMSNPANG--NSTTGTSTRITATLGGVSGYADLTVA PSLTISQIDPHTP 817

QY 302 PLVNLGLPQSFANFVAGIQPLTLDGIGKSOQNM-----WAGAPASAAATGTDIGKLP 356

DB 818 SVANGLTQ-----NETATGV-----YSDGNSQNLTDSTVWASSNPATVATSNAGTNGKAT 868

QY 357 PIQTSSGNSTARNGSSSTRYSQDGYPGDLVDVTITSEGLKQKYSNOVDVFNIPLA 416

DB 869 TLQGTGNSIASLGATTS-----DPSVLTVT-----NATLTSITIAPTS 907

QY 417 RFTSEDGLRRE-----GNHYSATLDSGGPGLPCTSNVGL 454

DB 908 SFNIAKGLNQDFVATGYTVDGSRDLTQVTVWNSNTSTATISNANGTQGRMAAVDTGST 967

QY 455 SVN-----QLETNSVDMREVMNMI IQRGFQWNSKSVTTADTML 494

DB 968 NISASLGTYTQTTNVTVTSVLNSIQVSPADISVAKGNTKAYTAI 1013

RESULT 31

US-09-881-752A-150

Sequence 150, Application US/09881752A

Patent No. US20020115078A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold

APPLICANT: Al-Garawi, Amal

APPLICANT: Miller, Charles

APPLICANT: Tomb, Jean-Francois

APPLICANT: Comen, Raymond P.

TITLE OF INVENTION: Identification of polynucleotides

TITLE OF INVENTION: Genome

FILE REFERENCE: 06132/041002

CURRENT APPLICATION NUMBER: US/09/881,752A

CURRENT FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 08/833,457

PRIOR FILING DATE: 1997-04-01

NUMBER OF SEQ ID NOS: 370

SOFTWARE: FastSeq For Windows Version 4.0

SEQ ID NO 150

LENGTH: 1230

TYPE: PRT

ORGANISM: Helicobacter pylori

US-09-881-752A-150

Query Match 6.2%; Score 160; DB 9; Length 1230;

Best Local Similarity 21.6%; Pred. No. 0.0011;

Matches 114; Conservative 71; Mismatches 182; Indels 160; Gaps 25;

QY 20 GTVSNNIANANTIGYKQQQVV--FQDLFSQDLAIGS-TGSGPNQAGMAQVGSVRTIPT 76

DB 542 GTTINTQAKSNAPKLKAMVVVNNEEBAAKTANLAQSSGTTTQSPNSTVNGA-----LNTVL- 596

QY 77 QGAPEPNSVTDLAIGKGFQVLTEDKVHYTRAGNERFT-QDGFLNDSGFTLMGSRIS 135

DB 597 -----QNVSNFQQSISQNAFQWNSNIQAWANAIY 625

QY 136 NPNKIKETLEP-----IQLDENDPTVAKS-PAKTSTALNAVVNIGDSTDKTQSBANP 187

DB 626 NTNCSQSQEMTPNNQDLRIQLRANFYQLINTINQVPTDMNALINQSQQTQOTSGSASN 685

QY 188 YFALLSWKNGTTPPISTNSYVAQPMRVYDQGNSHDITVYFPCAPSSTGSKTFEYLVA 247

DB 686 NNACASGMSG-----SNGNWCYQQ-----WSDSKAYYSGLQSALGYQT----- 723

QY 248 MNPSEDGSAASGTDAGLLMSGT--MTFSSNGELK-----NMTAFTPTGSGATK 293

DB 724 -----QATTQSGSGNGNSITVYVQIITLTSNGLLNQIITNLKSVNNGNGASGTSGSGTS 778

QY 294 DLNAWQPAFLVNLGLPQSFANFVAGIQPLTLDGIGKSOQNMWAGAPASAAATGTDIGKLP 353

DB 779 QIN-----TAYQMLT-----DASDGKLG- 797

QY 354 SMPLOTSSG-----NSTARNSSSTR-RYSQDGYPGDLVDVTITSEGLKQKYSN 404

DB 798 -----YSSSSGNGNGVTPCNST--NGSNKTSNCCYEPNKKQNTATATATTSNLSQKYND 851

QY 405 SOVDVFNIPLARFTS---EDGLR--EGNNHYSATLDS--GGPEFGLPCTSNVGL--S 455

DB 852 AQKI--ANI IASSGNKGVENGKQFPEALKNSSSLSNLCCNGSSGSSGTTCSGWLNL 909

QY 456 VNQLTSNVDMREVMNMI-----IQRGF-QWNSKSVTTADTMLQKAL 498

DB 910 LGAIPTGVSDTNNLLNLTTEFIKTAGFIQNNDSVSTSLTSAFOAI 956

RESULT 32

US-10-282-122A-5893

Sequence 5893, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

```
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58893
; LENGTH: 1230
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-282-122A-58893

Query Match
Best Local Similarity 6.2%; Score 160; DB 15; Length 1230;
Matches 114; Conservative 71; Mismatches 182; Indels 160; Gaps 25;

QY 20 GTVSNNIANTIGYKQOVV--FDLFSQDLAIGS-TGSGPQAGAGQVGSVRTFT 76
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
542 GTTNTQAKSNAPLKMVMVNEEAKTANLAQSSGTTTQSPNSTVNGA---LNTVL- 596
QY 77 QCAPEGNSVTDLAIGGKGFQVTELDKVVHTRAGNFRFT-QDGLNDPSGFTLMGSRIS 135
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
597 -----QVNSNFQOSIQNAFQCESNIOAWANAIY 625
QY 136 NNPNIKKTLEP-----IQDLFNDPTVAKS-PAKTSTALNAVNLGSDTDKTSQSEANP 187
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
626 NTNGSQSSEMTFNNQDLRIQLRANFVQLINTINQVPTDMNLINQSQQTQTSQGSAN 685
QY 188 YFALLESKMGNTPTSTNSYSYAPMRVYDQGNSHDITVYFDGAPSPSTGSKTFEYLVA 247
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
696 NNACASGMSG-----SNGNWCYQQ-----WSDSKAYVYSGLSQALGYQT----- 723
QY 248 MNPSEDGSAAGTDSAGLLMSGT--WTFSSNGELK-----NMTAFTPTGSATK 293
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
724 -----QATQSGSGNGSIITVVOQITLISNGLLNQIITLKVNGNGGASGTGSGNGTS 778
QY 294 DLNAWQAPLVNGLPQPSANFVGAGIQPLTLDFFGIKSQQNMWAGAPASAAAICTDICKLP 353
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
779 QIN-----TAYQMLT-----DASDGKLT----- 797
QY 354 SAMPLOTSSG-----NSTARGSSSTR-RYSQDGYPOGDLVDVTTITSEKLGKYSN 404
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
798 -----YSSSSGNGNYTCNST--NGSNKTSGNCCYEPNKOONATTATATDSNLQKYND 851
QY 405 SQVVDYFNIPARFTS---EDGLRR--EGNNHYSATLDS--GGPEFGLPQTSNYGKL--S 455
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
852 AQXI--ANIIASSGNKGVENGLKQPEALKNNSSLSNLCCNGSSGSGSTTCSGHLNLI 909
QY 456 VNQLETNSVNSREVMNMI---IIQRF--QMSKSVTTADTMDLQAL 498
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
910 LGAIFTNGVSDTNNLINLITFEIKTAGFIQNNDSSVSTSLTSAFOAI 956

RESULT 33
US-10-282-122A-49147
; Sequence 49147, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Olsen, Karl
; APPLICANT: Zyskind, Judith
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49147
; LENGTH: 3073
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49147

Query Match
Best Local Similarity 6.1%; Score 156.5; DB 15; Length 3073;
Matches 139; Conservative 73; Mismatches 228; Indels 187; Gaps 32;

QY 4 SLFTGAT-----GKTHSTGLGTVS--NNIANANTIGYKQOVVFDLFSQD-----LAIG 52
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2164 SLLLCAPVTNVIYAVSQNVTCGGSTASNDLNAIAG-----PLAAASGVGALAVG 2214
QY 53 S---TGSGPQAGAGQVGSVR--TIQGAPEFGNSVTDLAIG-----GKFFQVTLDEK 104
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2215 AGSTAGSDASTAVGTGAAGVGSVNSTAIQYASVGVNSANSLSLAIGYNSRAQLNSLAIGTE 2274
QY 105 VHYTRAGNFRFTQDGLNDPSGFTLMGSRISNNPNKKTLEPIQLDFN---DPTVAKSP 161
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2275 ASATSAGSIAIGYCAFLN--PSATNSMALGL-----NSSVSAANAIVAICYNVADRANVS- 2328
QY 162 AKTSTALNAVNLGSDTDKTKQS--BANFYFALLEBWKNGTNP----- 201
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2329 VGSKKQSQSIINVAAGTANTDAVNLGQMNAAINAVAGGSPNAVYVDTSAHMSVTLGSAG 2388
QY 202 -PISTSNYSYAQPMRVYDQGNSHDITV---YFDGAPSTGSKTPEYLVAMNPSEDSG- 255
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2389 TPVKVSN-----VANGVANDAVNVAQLTAMGTGTTINSGVVTSNF--VAYDDTTKGS 2439
QY 256 ---AASGTD---SAGLL-----MEGTWTFSSNGELKNWTAFTTP--TGSATKDLNNAWQ 299
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2440 TLKGASGTTITNVKAGSLTANSTDAINGOLYOTNANVANLANVANITGNVTNTVN--- 2496
QY 300 PAPIVN--GLPQFSAN----- 313
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2497 --NIWNGGGIKYFHANSTLADSSATGTDSDVAIGAANATANSVALGANSVAGRANAVSV 2554
QY 314 -FVGAGIQPLTLDPGIKS-----QQNMWAGAPASAAAICTD-----IGKL 352
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 2555 GAVGAERQIINAVANTSTDAVNLSQLOAMGANVNSGVVTAFAVDDSTKGVTLGGV 2614
QY 353 PNMPI-----QTSSGNSTARNSSSTRYSQDGYPOGDLVDV-----TITSEKLOG 400
Db 2615 GSSMPVTLNVAEGQVTSKQALNSQLYGTANSVASALGCTSSVSGNSGNTVAAFSLDG 2674
QY 401 KYNSN--QVVDYFNIPLARFTSEDLREGNNHYSATLDSGGPEFGLP-----GTSN 450
Db 2675 KYNSVATMDALNAKIAT--GSTDGVVDTSAHKNLTL--GGVNATTPVTVANVAATS 2731
QY 451 YKGLSVNOLETS--NVDMSREVMNMII 475
Db 2732 DOAVNLAOLKAAGLVNDSGNVTSFV 2758
RESULT 34
US-10-369-493-13825
; Sequence 13825, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13825
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13825

Query Match 5.9%; Score 153.5; DB 14; Length 679;
Best Local Similarity 20.3%; Pred. No. 0.0015;
Matches 127; Conservative 69; Mismatches 208; Indels 221; Gaps 30;
QY 5 LFIGATGKTHSTGLGTVSNNIANTIGYKQQQWFDLFSQDLAIGSTGSGPNQA-- 62
Db 2 LNIWMSGLAASSSSSLAVTGNNIANVDITAGYSRQTV-----QGTSSSIQ 45
QY 63 -----GNGAQVGSVRTFTQGAPEPGNSVTDLAIGKGP-FQVTLKDYHYTAGNFRF 115
Db 46 YGNVFIGTTLADVRRVYNSYLESQLRTATSLNSEAAAFGAQATALD-----GSLSD 98
QY 116 TDGFLNDPSGF--TLMGSRISNNPNIKETL-----EPIQLDFNDPTVAKSPAKTSTALN 169
Db 99 TWTGLTGLVKFTSMQGVSTSDTDSRSQSVLTGAQALTSREN--ALAKQMDQXATLN 156
QY 170 AVVNLGSDTDKTOSEANPYFALLSKWNGTPTTSTNSYVAQPMRYD----- 218
Db 157 G--NLSDMASQVKNLATS-IATLNQKTGE-----ISTSG--GQPNLLDSRNEAVRQLSE 206
QY 219 -----QQGNSHDI-----TVYFDGAPS-----STGSKTFEYLYVA 247
Db 207 LTGAQVVERGTFDIYIGSGQPLVIGNTNTLSTVPLKGPFSRMGIQMDRSGSTIITSA 266
QY 248 MNPSEGG-----SAASGTSAG----- 264
Db 267 MTGGEIGLLTYRKEVLDPALNELGRVALVADQINRQQAQGDIDKNGDFGAALFNNNSA 326
QY 265 -----LLMSGT-----MTFSSNGEL-----KNMTAFTGTSATKDLNAWQ-- 299
Db 327 ALISQRTAQSGNSAGSNLDVTIKDTGLTTSQVYVFTTSATNYTVKRSDGTDMSFSF 386
QY 300 ---PAPLVNGLPQSFANFVGAGIOPFLIDFGIKSQQNNWAGAPASAAA-----IGTDIGK 351

Db 387 TTTTPPVVDGF-----TLALNGALASGDTFKVTTRNAASSIQVLTDPKK 433
QY 352 LPSMPTQTSNGNSTARNSSSTRYSQDGYPOGDLVDV-----TITSEKLOG--KYNSN 405
Db 434 IAAAGPL--TGVASANNST-----YTQPTLTDVVDIYNPASQABLQNALKYSTP 481
QY 406 QVVDYFNIPLARFTSEDLREGNNHYSATLDSGGPEFG-----LPGTSNYKLSVNOLET 461
Db 482 VKLVP-----GATTSQSQTYN-MVDAKGATIGSGVIVPGQANTLNLIKGMVDS 528
QY 462 SNVDMRSREVMNMII--IQRFQMN 484
Db 529 TGAAPV--MDTVPVNPVKITFTVQT 550
RESULT 35
US-10-335-977-4779
; Sequence 4779, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4779:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...269
; SEQUENCE DESCRIPTION: SEQ ID NO: 4779:
US-10-335-977-4779

Query Match 5.9%; Score 151.5; DB 15; Length 269;
Best Local Similarity 25.9%; Pred. No. 0.00057;
Matches 58; Conservative 39; Mismatches 90; Indels 37; Gaps 8;
QY 6 FIGATG-MKCHSTGLGTVSNNIANTIGYKQQQVVFQD-----LFSQDLAI 51
Db 5 YTAATGAMATQFNRLDTSNNLANLNINFGRRDDAITGDFLRLYQYRQPLEDOTKAS 64

QY	52	GTSQGPNQAGMGOVGSVETIIPTQAFEPGNSVTDLAIGKG-FFQVTELEDKHVHTRA	110
Db	65	AKYLNRNLNRPILISEIYDTRSL--CAFEGTNPLDFALTSNLPFAQTQHEGVAYTRD	121
QY	111	GNFRPTQGFPLNDPSGFTLMGSRISNN-----PNIKKETLPIQLDFNPOTVAKSPA	162
Db	122	GHFSYDKGGLFVLTLNGPKVL-SRSGLNKEGGIMLPNAEIEVPQNGGITFRD-----NEA	175
QY	163	KTSTALNAVNLGDSTDXTQSEANPYFALLEWSKNGCTPPISTS	206
Db	176	QIQAGALAVSFSPFNKLKIGONLY-----TYQGEVGHVOVS	214

RESULT 36

US-10-282-122A-64364
; Sequence 64364, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```

Query Match          5.9%; Score 151.5; DB 15; Length 2204;
Best Local Similarity 20.2%; Pred. No. 0.012;
Matches 119; Conservative 64; Mismatches 220; Indels 185; Gaps 24;

yy      3  GSLFICATGMKTHSTGLGFSVNSNIANANTIGYKQOVVFQDLFQDLAGSTG-----55
bb      1458  GNIGIGLATG--TCQIGFSGFSNSHIGLFGNSGSGDNVGFNFGSTGTGVGIGITGTANFGIA 1515
yy      56  SGQPNQAGCAQGVGVRIIFTQGAEPGNSVTDLAGCK-----GPFQV-----99

```

```

1516 NSGSFNTGLG-NTGSTNT-----GLFNPGNVNTGVNGTSGINTGSGINTGSGNTGTSF 1570
100 TLED-----KVHYTRAGNPRF-----TQDGF-----120
1571 NLGDHNTGSPNSGDYNTGYFNACDYGTVGANTGVNTGAFISGYSNGFWRGDYQGLG 1630
121 -----LNDPSGFTLMGSRISNNPN-----IKKTTLP 147
1631 LSTTIPIEIPYRDLVSVIDIPITGVVATTNPSFTIPGQIRVLLGPAALVNMEMIG 1690
148 IQDLDFNDPTVAKSPAKTSTALANAVNLGSDTKDTGEANPYFALLESWKNGTPTPISTN 207
1691 ITIDVNOVAIDSPIOQTSMVGTGCGFPIGISIGTGGF-----GNSF-----1736
208 YSAQPMRYVDQGNSHDITVFDGAPSTGSKTTEYLVAMNPBDDGSAAG-TDSAGLL 266
1737 ---TGSSGFFHTGAGVSGFGNGAGNMSGNCF-----GAGNSOFFNAGGIG 1782
267 MSGTMTFSS-NGBELKMTAFTPTGSATKDLNAWQAPLVNLPQPSANFVGAGIQ---PL 322
1783 NSGLNFGALQSLANLGNITSGVYNTSTLDIATPA-FGSGIANTGANLAGLFLONTGNL 1841
323 TLDPGIKSQONWAGAPASAAIGTIDIGLPSMPPIOTSSGNSGTARNSSSTRY-----377
1842 TLNFGVANQGLNAG-----IGNLGSWNTGFVNTGSDNLGIGNLGLDNLFNGGVNI 1890
378 ---SQDGPQGDLDVTTITSEGLKQKYSN--SQVDF-----YNTPLARFTSED-----422
1891 GGNNTGIANTGTFDGLANLGSYNTGLANLGDNDNLGFGNAGSYNLGFANFGSDNLGFANT 1950
423 -----GLRRBGNHYS-----ATLDSGGPEFGL--PGFSNYG 452
1951 GSYNTGFANTGNNTGVGLTGQIGIGISLNSGSGNNIGLNFNSGSGNIG 1998

```

RESULT 37

```

US-10-193-764-61
/ Sequence 61, Application US/10193764
/ Publication No. US20030133943A1
/ GENERAL INFORMATION:
/ APPLICANT: Loosmore, Sheena M.
/ APPLICANT: Yang, Yan-Ping
/ APPLICANT: Klein, Michael H.
/ TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
/ TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
/ FILE REFERENCE: 1038-1239MIS
/ CURRENT APPLICATION NUMBER: US/10/193,764
/ CURRENT FILING DATE: 2002-07-12
/ PRIOR APPLICATION NUMBER: 09/167,568
/ PRIOR FILING DATE: 1998-10-07
/ NUMBER OF SEQ ID NOS: 91
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 61
/ LENGTH: 1180
/ TYPE: PRT
/ ORGANISM: Haemophilus influenzae
US-10-193-764-61

```

```
Query Match      5.8%; Score 149.5; DB 14; Length 1180;
Best Local Similarity 20.2%; Pred. No. 0.0072;
Matches 113; Conservative 88; Mismatches 212; Indels 147; Gaps 22;

QY          30 NTICYKQQVVFQDLFSQDLSATGTSQGPNCAAGVGSVRIFQTQGAPEPGNSVTDL 89
              |||:::||:||||:||||:||||:||||:||||:||||:||||:
Db          606 NISGFNKAKEIVAKD--SSNLITGNSDDSGNT-----SAKTIV-----TFNNVKOS 647

QY          90 AIGSGKFFQVLTEDKVHYFRAGNFETQGFGLNDPSGFTLNGSISNNPIKKETLEPIQ 149
              |||:::||:||||:||||:||||:||||:||||:||||:||||:
Db          648 KISADG-HKYTLNSKVK-TLSDNDNTEGGSDNN-TGLTITAKDEVVNHNIT-----696

QY          150 LDFNDTVAKSPA-----KSTALNAVNI LGSDTKTQS-----EANPYFALLESW---195
              |||:::||:||||:||||:||||:||||:||||:||||:||||:
Db          697 ---SHKTVNVSAAANGIITKTGTITINATAGNVETAHTGTGIQGGIESKPQGSVITAVAGDGT 753
```

QY 196 -----KGN-----GTPPISTNSYSYAPQMRVYDQGN----- 222
Db 754 LAVGNISGNATVTANSALTTLAGSTIKGTESITSS-----QSGNIGKISG 802
QY 223 -----SHDITVYFDGAPSSGSKTTEYLVMNPSDEGSAASG-----TDSAG 264
Db 803 KTVNKAATNSLTQADSKIEATEG-----EANTVSKTSIIGGTISGGTVEVTATEGLTTQAG 859
QY 265 LLMSGTWTFFSSNGELKNMTAFTPTG-----SATKDL-----NMQPAPLVN 305
Db 860 STITGESVTTSSQSGNIGMISGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTD 919
QY 306 GLPQFSANFVAGAGIQLTLDGFIKSQQNMWAGAPASAAAIGTDIGKLPSPMPIQTSSGNS 365
Db 920 GTISGNTVNTANTGDLTVEDEAK-----IDATGGAATLTATSGKL-----TTKASS 966
QY 366 TARGSSSTRYSQDGYPOQDL--VDVTITSEKLOGKYSNQVDFVNIPLARFTSEBG 423
Db 967 SITSAANQVNLAKDGSIGGNINAANVTLTNTGALTTVKSSINANSGLTVINAKDAELN 1026
QY 424 LRREGNNHYSATLDSGGPEGLPGTSGNYKLSVNOLETNSVD--MSREVMNMIILQRFQM 482
Db 1027 GEASGNTVNTANTGDLTVEDEAK-----IDATGGAATLTATSGKL-----TTKASS 966
QY 483 NKSQVTTADTLMQKALELKR 502
Db 1086 DVKYIQGIASVDEVIEAKR 1105
RESULT 38
US-10-193-764-59
; Sequence 59, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; FILE REFERENCE: 1038-1239MIS
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US/10/193,764
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 59
; LENGTH: 1188
; TYPE: PR1
; ORGANISM: Haemophilus influenzae
US-10-193-764-59
Query Match 5.8%; Score 149.5; DB 14; Length 1188;
Best Local Similarity 20.2%; Pred. No. 0.0073;
Matches 113; Conservative 88; Mismatches 212; Indels 147; Gaps 22;
QY 30 NTIGYKQOVVFDLFSODLAIGSTGSGQGNQAGWAGVGSVRTIFTQGAPEGNSVTDL 89
Db 614 NISGFKAKIVAKD--SSNLITGSDSDSGNT-----SAKTV-----TFNVKDS 655
QY 90 AIGKGFFQVTLDEKHYTRAGNFRFTQDGLNDPSPGFTLMGSRISNNPNIKKLEPIQ 149
Db 656 KISADG-HKVTLSNVK--TLDNDNTEGSDNN--TGLTITAKDVEVNNIT----- 704
QY 150 LDFNDPTVAKSPA-----KTSTALNAVNLGSDTDKTS-----EANPYFALLESW--- 195
Db 705 ---SHKTVNVAANGGIITKTGTINATAGNVEITAHTGSGQGGIESKPSGVSITVAGGDT 761
QY 196 -----KGN-----GTPPISTNSYSYAPQMRVYDQGN----- 222
Db 762 LAVGNISGNATVTANSALTTLAGSTIKGTESITSS-----QSGNIGKISG 810

QY 223 -----SHDITVYFDGAPSSGSKTTEYLVMNPSDEGSAASG-----TDSAG 264
Db 811 KTVNKAATNSLTQADSKIEATEG-----EANTVSKTSIIGGTISGGTVEVTATEGLTTQAG 867
QY 265 LLMSGTWTFFSSNGELKNMTAFTPTG-----SATKDL-----NMQPAPLVN 305
Db 868 STITGESVTTSSQSGNIGMISGKVEVSATKDLITKSGSEIKATAGEVNVTSATGTD 927
QY 306 GLPQFSANFVAGAGIQLTLDGFIKSQQNMWAGAPASAAAIGTDIGKLPSPMPIQTSSGNS 365
Db 928 GTISGNTVNTANTGDLTVEDEAK-----IDATGGAATLTATSGKL-----TTKASS 974
QY 366 TARGSSSTRYSQDGYPOQDL--VDVTITSEKLOGKYSNQVDFVNIPLARFTSEBG 423
Db 975 SITSAANQVNLAKDGSIGGNINAANVTLTNTGALTTVKSSINANSGLTVINAKDAELN 1034
QY 424 LRREGNNHYSATLDSGGPEGLPGTSGNYKLSVNOLETNSVD--MSREVMNMIILQRFQM 482
Db 1035 GEASGNTVNTANTGDLTVEDEAK-----IDATGGAATLTATSGKL-----TTKASS 974
QY 483 NKSQVTTADTLMQKALELKR 502
Db 1094 DVKYIQGIASVDEVIEAKR 1113
RESULT 39
US-10-335-977-4780
; Sequence 4780, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4780:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...273
; SEQUENCE DESCRIPTION: SEQ ID NO: 4780:

Db 171 LGFDKSSSTANPPFFGMEAWDGTATPPLGNAAGYTQTMVYDAAGDTHLTIREFDNA 230
QY 234 PSSTGSKTEYIVANVPSEDSAGAGTDSAGLLMSGTWTFSSNGELKNMTAFTPTGSATK 293
Db 231 TTAGGRKRVYEFVLVTPPDEGGAAGSAGAGLLMSGTLLQFGSSGQLODIMAFTPSGGDPK 290
QY 294 DLNMQPAPL-VNGLPQFSANFVGAGIQPLTLDFGIKSQQNNWAGAPASAAAIGTDIGKL 352
Db 291 DLANNVPATLDASGRPQFNATFAOQGAQTVGLDLGIT--TANNAPASAAAGVGLDPTLL 348
QY 353 PSWMIQITSSNGSTARNSSSTRYSODGYPOGDLVDVITSEGLQKYSNSQVDFYN 412
Db 349 GGATTPKLAATSTTAAYKSSSTYKQDGYAAGVLMNLEAVTDGIVSGKYSNGQSOLF 408
QY 413 IPLARFTSEDLGRREGNNHYSATLDSGGPEFGLPGTNSYKLSVNQLETSNVDMREMN 472
Db 409 VPIFTSEDLGRSEGMNHSATTESGAQEGKADTENYKLGKMGKSLQSNVDMAREMN 468
QY 473 MIIORGQFMSKSVTTADTMLOKALELKR 502
Db 469 MIVTORGQFMSKSVTTADTMLOKALELKR 498

RESULT 2

AA94790 PRELIMINARY; PRT; 498 AA.
AC AAS94790;
DT 27-APR-2004 (TREMELrel. 27, Created)
DT 27-APR-2004 (TREMELrel. 27, Last sequence update)
DT 11-MAY-2004 (TREMELrel. 27, Last annotation update)
DE Flagella basal body rod domain protein.
GN DVU0307.
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15077118; DOI=10.1038/nbt959;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N., Methe B.A., Brinkac L.M.,
RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Daviden T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
Desulfovibrio vulgaris Hildenborough.";
RL Nat. Biotechnol. 22:554-559(2004).
DR EMBL; AE017310; AAS94790.1; -.
DR TIGR; DVU0307; -.
DR Flagella.
SQ SEQUENCE 498 AA; 51954 MW; CDE6B383774E8807 CRC64;

Query Match 47.8%; Score 1235; DB 2; Length 498;
Best Local Similarity 49.6%; Pred. No. 5.1e-67;
Matches 253; Conservative 74; Mismatches 163; Indels 20; Gaps 5;

QY 1 MMGLSFGATGKMTSTGLGTVSNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGQGN 60
Db 1 MMGLYTGATGMRTQAEGMNVVGNLNVNTIGFKQSMMLYQDLMSQTEPTGSAVYSGIS 60
QY 61 QAGMGAQVGSVRTFTTQGAPEPGNSVTDLAIGKGGFFQVTLKDVHYTRAGNFRFTQDGF 120
Db 61 QVGLGARGVDVRLRSQALLAGSDTDFALSGKGFQVTSQGDTHYTRAGNFRFNKQDQ 120
QY 121 LNDPSGFTLMSRISNNPNIKETLEPIQLDFNDPTVAKS-----PAKTSTALNAVYN 173
Db 121 LVDNPNGLMHAITG-----ETEGFLAA-----VTLVKDADGRLSNPAKATGGMTSVFN 170
QY 174 LGSDTKTQSEANPYFALLESWKNGTTPPTISNYSYQAQPMRVYDQGNSHDITVYDFCA 233

Db 171 LGFDKSSSTANPPFFGMEAWDGTATPPLGNAAGYTQTMVYDAAGDTHLTIREFDNA 230
QY 234 PSSTGSKTEYIVANVPSEDSAGAGTDSAGLLMSGTWTFSSNGELKNMTAFTPTGSATK 293
Db 231 TTAGGRKRVYEFVLVTPPDEGGAAGSAGAGLLMSGTLLQFGSSGQLODIMAFTPSGGDPK 290
QY 294 DLNMQPAPL-VNGLPQFSANFVGAGIQPLTLDFGIKSQQNNWAGAPASAAAIGTDIGKL 352
Db 291 DLANNVPATLDASGRPQFNATFAOQGAQTVGLDLGIT--TANNAPASAAAGVGLDPTLL 348
QY 353 PSWMIQITSSNGSTARNSSSTRYSODGYPOGDLVDVITSEGLQKYSNSQVDFYN 412
Db 349 GGATTPKLAATSTTAAYKSSSTYKQDGYAAGVLMNLEAVTDGIVSGKYSNGQSOLF 408
QY 413 IPLARFTSEDLGRREGNNHYSATLDSGGPEFGLPGTNSYKLSVNQLETSNVDMREMN 472
Db 409 VPIFTSEDLGRSEGMNHSATTESGAQEGKADTENYKLGKMGKSLQSNVDMAREMN 468
QY 473 MIIORGQFMSKSVTTADTMLOKALELKR 502
Db 469 MIVTORGQFMSKSVTTADTMLOKALELKR 498

RESULT 3

Q72C41 PRELIMINARY; PRT; 564 AA.
AC Q72C41;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE Flagellar hook protein FlgB.
GN Name=flgB; Order=sluocNames=DVU1443;
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15077118; DOI=10.1038/nbt959;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N.L., Methe B.A., Brinkac L.M.,
RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Daviden T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
Desulfovibrio vulgaris Hildenborough.";
RL Nat. Biotechnol. 22:554-559(2004).
DR EMBL; AE017314; AAS95921.1; -.
DR TIGR; DVU1443; -.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR001444; Flag_bb_rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF00460; Flg_bb_rod; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
KW Complete proteome; Flagellum.
SQ SEQUENCE 564 AA; 60342 MW; 50DC0B4F6D7A4DBE CRC64;

Query Match 37.2%; Score 961; DB 2; Length 564;
Best Local Similarity 39.0%; Pred. No. 2.9e-50;
Matches 223; Conservative 85; Mismatches 184; Indels 80; Gaps 17;

QY 1 MMGLSFGATGKMTSTGLGTVSNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGQGN 60
Db 3 LTASWTGVSGLLAGRMNVNLNNIANVNTVFGKSRMDFEFLNQD-----TVSAAGVT 58
QY 61 QAGMGAQVGSVRTFTTQGAPEPGNSVTDLAIGKGGFFQVTLKDVHYTRAGNFRFTQD 118
Db 59 QVGRGVSIGALFGDYSQAGAFQTTNESTDLATQGRGFFSVKPKGTEDTYTTRAGNFRFDAD 118


```

QY 119 GFLNDPSGFTLMGSRISNNPN-----IKKTELEPIQLDFNDPTVAKSPAK 163
DB 119 GYLVDPHGVVLOQWAIERSQNSLTSSAVTATSTTSKIGSGV-PVDIKL-DGFTAEPOHT 176
QY 164 TSTALNAVNLGSDTKQSEANPYFALLESKWG-----NGT-PPISTSNYSYAPQMRVYD 218
DB 177 QNITLNVLDSPGNDKSSSTNPFSLFETWNGQNLPTGTQPALAQSAFAYQSTIKVYD 236
QY 219 OQNSHDITVYFDGA-PSS-----TGSKTFEYLVMNPSED-----GSAASGTDASGLLM 267
DB 237 EAGTAHTLTVYFDQVDPDSVTNEPNRQKWEYIVTMDPAEDKRVIACTAMNTTAAAGLLM 296
QY 268 SGTMTFSSNGELKNMTATPTGSGAT-----KDLNMQAPL-VNG 306
DB 297 TGLTFDTGQLVDQATFTPGQYDTTPPNNEPTNPGPPVTPPDLVNMQPTQSSNG 356
QY 307 LPQFSANFVG-----AGIOPLTLDGFIKSSQ--NNWAGAPASAAAIGTDIGKLP 353
DB 357 LPMVAFNSGLTSSVVGSPTAQNFMELDLGLASTNATTPWTSTP-NAAAIGTDASLLP 415
QY 354 SMPTQTSNGSTARNSSSTRYSQDGYPOGDLVDVTTITSEKLOGKYSNSQVVDPYNI 413
DB 416 GLTSSQRPQSATTSVAGSSSTQFQKQDGYTFGLQNTIVDRDGMQKYSNGVTLDIYQV 475
QY 414 PLARTSBDGRRGNHYSATLDSGGPEFGLPQTSN---YKGLSVNQLTSNVDMMSREM 470
DB 476 TLVDFTSKNLRREGNLFSGTRDSDGP---LPGPANSNGLGAISSNLSQSNVDLAREF 532
QY 471 VNMIIQRFQFQNSKSVTTADTMLOKALELKR 502
DB 533 VEMITQRFQFQNSKSVTTTDMLEVVVNMKR 564

RESULT 4
AAS95921 PRELIMINARY; PRT; 564 AA.
AC AAS95921;
DT 27-APR-2004 (TremBLrel. 27, Created)
DT 27-APR-2004 (TremBLrel. 27, Last sequence update)
DE 11-MAY-2004 (TremBLrel. 27, Last annotation update)
DE Flagellar hook protein FlgE.
GN FLGE OR DVU1443.
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobivionales;
OC Desulfobivriaceae; Desulfobivrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed15077118;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N., Methe B.A., Brinkac L.M.,
RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
RL Desulfovibrio vulgaris Hildenborough."
RL Nat. Biotechnol. 22:554-559(2004).
DR EMBL; A2017314; AAS95921.1; -.
DR TIGR; DVU1443; -.
KW Flagella.
SQ SEQUENCE 564 AA; 60542 MW; 50DC0B4F6D7A4DBE CRC64;

Query Match 37.2%; Score 961; DB 2; Length 564;
Best Local Similarity 39.0%; Pred. No. 2.9e-50;
Matches 223; Conservative 85; Mismatches 184; Indels 80; Gaps 17;

QY 1 MMGSLPIGATGMKTHSTGLGTSSNNIANANTIGKQQVQFDLPFSQDLAIGSTGSGCPN 60
DB 3 LTASMTGVSGLLAHGERANVLGNINANTVVGFKGRMDFEDFLNQD----TYSAGVNT 58

```

```

QY 61 QAGMGAQVGSVRTITFGAFEPGNSVTDLAIGGKGFQVTL--DKVHYTRAGNFRFTQD 118
DB 59 QVGRGVSGIAGFYGSQAFQCTTNESTDLAQGRGPFSPKGTEDTYTTRAGNFRFDAD 118
QY 119 GFLNDPSGFTLMGSRISNNPN-----IKKTELEPIQLDFNDPTVAKSPAK 163
DB 119 GYLVDPHGVVLOQWAIERSQNSLTSSAVTATSTTSKIGSGV-PVDIKL-DGFTAEPOHT 176
QY 164 TSTALNAVNLGSDTKQSEANPYFALLESKWG-----NGT-PPISTSNYSYAPQMRVYD 218
DB 177 QNITLNVLDSPGNDKSSSTNPFSLFETWNGQNLPTGTQPALAQSAFAYQSTIKVYD 236
QY 219 OQNSHDITVYFDGA-PSS-----TGSKTFEYLVMNPSED-----GSAASGTDASGLLM 267
DB 237 EAGTAHTLTVYFDQVDPDSVTNEPNRQKWEYIVTMDPAEDKRVIACTAMNTTAAAGLLM 296
QY 268 SGTMTFSSNGELKNMTATPTGSGAT-----KDLNMQAPL-VNG 306
DB 297 TGLTFDTGQLVDQATFTPGQYDTTPPNNEPTNPGPPVTPPDLVNMQPTQSSNG 356
QY 307 LPQFSANFVG-----AGIOPLTLDGFIKSSQ--NNWAGAPASAAAIGTDIGKLP 353
DB 357 LPMVAFNSGLTSSVVGSPTAQNFMELDLGLASTNATTPWTSTP-NAAAIGTDASLLP 415
QY 354 SMPTQTSNGSTARNSSSTRYSQDGYPOGDLVDVTTITSEKLOGKYSNSQVVDPYNI 413
DB 416 GLTSSQRPQSATTSVAGSSSTQFQKQDGYTFGLQNTIVDRDGMQKYSNGVTLDIYQV 475
QY 414 PLARTSBDGRRGNHYSATLDSGGPEFGLPQTSN---YKGLSVNQLTSNVDMMSREM 470
DB 476 TLVDFTSKNLRREGNLFSGTRDSDGP---LPGPANSNGLGAISSNLSQSNVDLAREF 532
QY 471 VNMIIQRFQFQNSKSVTTADTMLOKALELKR 502
DB 533 VEMITQRFQFQNSKSVTTTDMLEVVVNMKR 564

RESULT 5
Q74G30 PRELIMINARY; PRT; 419 AA.
ID Q74G30
AC Q74G30;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Flagellar hook protein FlgE.
GN Name=flgE; OR=Names=GSU0419;
OS Geobacter sulfurreducens.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;
OC Geobacteraceae; Geobacter.
OX NCBI_TaxID=35554;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PCA / ATCC 51573;
RX PubMed14671304; DOI=10.1126/science.1088727;
RA Methe B.A., Nelson K.E., Eisen J.A., Paulsen I.T., Nelson W.C.,
RA Heidelberg J.F., Wu M., Ward N.L., Beanan M.J., Dodson R.J.,
RA Madupu R., Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
RA Gwinn M.L., Kolonay J.F., Sullivan S.A., Haft D.H., Selengut J.,
RA Davidtsen T.M., Zafar N., White O., Tran B., Romero C., Forberger H.A.,
RA Weidman J.P., Khouri H.M., Feldblyum T.V., Utterback T.R.,
RA Van Aken S.E., Lovley D.R., Fraser C.M.;
RT "Genome of Geobacter sulfurreducens: metal reduction in subsurface
environments."
RL Science 302:1967-1969(2003).
DR EMBL; AEO17180; AAR33751.1; -.
DR TIGR; GSU0419; -.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR001444; Flag_bb rod.
DR InterPro; IPR002371; Flag_hookAPI.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF00460; Flag_bb rod; 1.

```


A Sockett R.E., Schuster S.C.;

RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
 RL genomic perspective.",
 DR Science 303:689-692 (2004).
 DR EMBL; BX842655; CAE78193.1; -.
 DR InterPro; IPR010930; DUF1078.
 DR InterPro; IPR011491; Flae.
 DR InterPro; IPR001444; Flae_bb_rod.
 DR Pfam; PF06429; DUF1078; 1.
 DR Pfam; PF07559; Flae; 1.
 DR Pfam; PF00460; Flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAELLA_BB_ROD; 1.
 KW Complete proteome.
 SQ SEQUENCE 427 AA; 45943 MW; C74DA82F2E0FC003 CRC64;

Query Match 22.4%; Score 579.5; DB 2; Length 427;
 Best Local Similarity 30.6%; Pred. No. 3.3e-27;
 Matches 160; Conservative 73; Mismatches 171; Indels 119; Gaps 16;
 QY 1 MMGSLFPGATGKTHSTGLTGVSNNTANANTIGYKQOVVFDLPSODLAIGSTGSGQPN 60
 DB 3 ILSSLYTVGSGMTAQGEALGVGDNIANANTIGFKASRAEFQDIISKNL-----KGIVG 58
 QY 61 QAGMGAGVGSVRTITQGAFFPGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGF 120
 DB 59 QIRGVKIGAVNPILSQGNIDATEKVTDLAISGDGYFKVKGSDGESYTRDGSFHFDRGY 118
 QY 121 L--NDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDST 178
 DB 119 LVTDN-----NQRVQGFSTDEKGNIVNKMTDIKPPR-ALIPAKATKELKLDNLDSRM 170
 QY 179 DKTQSEANPYFALLSWKNGNTPPISTSNYSYAOQPMRVYDQGNSHDITVYFDGAPSTG 238
 DB 171 EPTK-----KFPDAPYSTSHYSTG--VEMYDSQGNKHLVSFFP-----NKVN 211
 QY 239 SKTFEY--LVAMNPSDEGSAAGTDSAGL--LMSGTMTFSSNGELKNM-----TAFTTGS 290
 DB 212 DREWEFKGLV-----DGKEITGEEGKMSVAAGKLMFTVDGKLDSEQTSTNPNFKGG 265
 QY 291 ATKDLNMQAPLVNGLPQFSANFVGAGIQLPTLDFG--IKSQQNMWAGAPASAAAI 348
 DB 266 ALQD-----QQVKNLFGDAIK-----D 282
 QY 349 IGKLPMMPIQTSSGNSTARNSSSTRYS-----QDGYPOGLVDVTTITSEGLK 399
 DB 283 GKK-----GLDGTQYKGNKSLISWHQDGAAGTITGLSFNDEGTLT 324
 QY 400 GKYNSQVVDVFNPIPLAFTSEDLREGNNHYSATLDSGPEGLPGTSNYGKLSVNL 459
 DB 325 AVYSNGQANDLAQIALAKFENPEALFKVGNRLKESRDSGTASVGAAPGAGKGLFAKSL 384
 QY 460 ETSNVDMSREVMNMIIOQGFQMSKSVTTADTLMQKALELKR 502
 DB 385 ERSTVDLATEFVNMIQORGFQANAKITTTTDELLNEVIQLR 427

RESULT 10

CAE78193
 ID CAE78193 PRELIMINARY; PRT; 427 AA.
 AC CAE78193;
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE FlgE protein.
 GN FlgE OR B03395.
 OS Bdellovibrio bacteriovorus.
 OC Bacteria; Proteobacteria; Deitaproteobacteria; Bdellovibrionales;
 OC Bdellovibrionaceae; Bdellovibrio.
 OX NCBI_TaxID=959;
 RN [1]
 RL J. Bacteriol. 176:3631-3637 (1994).
 CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,
 RA Sockett R.E., Schuster S.C.;
 RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
 RL genomic perspective.",
 DR Science 303:689-692 (2004).
 DR EMBL; BX842655; CAE78193.1; -.
 SQ SEQUENCE 427 AA; 45943 MW; C74DA82F2E0FC003 CRC64;

Query Match 22.4%; Score 579.5; DB 2; Length 427;
 Best Local Similarity 30.6%; Pred. No. 3.3e-27;
 Matches 160; Conservative 73; Mismatches 171; Indels 119; Gaps 16;
 QY 1 MMGSLFPGATGKTHSTGLTGVSNNTANANTIGYKQOVVFDLPSODLAIGSTGSGQPN 60
 DB 3 ILSSLYTVGSGMTAQGEALGVGDNIANANTIGFKASRAEFQDIISKNL-----KGIVG 58
 QY 61 QAGMGAGVGSVRTITQGAFFPGNSVTDLAIGGKGFQVLTEDKVHYTRAGNFRFTQDGF 120
 DB 59 QIRGVKIGAVNPILSQGNIDATEKVTDLAISGDGYFKVKGSDGESYTRDGSFHFDRGY 118
 QY 121 L--NDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDST 178
 DB 119 LVTDN-----NQRVQGFSTDEKGNIVNKMTDIKPPR-ALIPAKATKELKLDNLDSRM 170
 QY 179 DKTQSEANPYFALLSWKNGNTPPISTSNYSYAOQPMRVYDQGNSHDITVYFDGAPSTG 238
 DB 171 EPTK-----KFPDAPYSTSHYSTG--VEMYDSQGNKHLVSFFP-----NKVN 211
 QY 239 SKTFEY--LVAMNPSDEGSAAGTDSAGL--LMSGTMTFSSNGELKNM-----TAFTTGS 290
 DB 212 DREWEFKGLV-----DGKEITGEEGKMSVAAGKLMFTVDGKLDSEQTSTNPNFKGG 265
 QY 291 ATKDLNMQAPLVNGLPQFSANFVGAGIQLPTLDFG--IKSQQNMWAGAPASAAAI 348
 DB 266 ALQD-----QQVKNLFGDAIK-----D 282
 QY 349 IGKLPMMPIQTSSGNSTARNSSSTRYS-----QDGYPOGLVDVTTITSEGLK 399
 DB 283 GKK-----GLDGTQYKGNKSLISWHQDGAAGTITGLSFNDEGTLT 324
 QY 400 GKYNSQVVDVFNPIPLAFTSEDLREGNNHYSATLDSGPEGLPGTSNYGKLSVNL 459
 DB 325 AVYSNGQANDLAQIALAKFENPEALFKVGNRLKESRDSGTASVGAAPGAGKGLFAKSL 384
 QY 460 ETSNVDMSREVMNMIIOQGFQMSKSVTTADTLMQKALELKR 502
 DB 385 ERSTVDLATEFVNMIQORGFQANAKITTTTDELLNEVIQLR 427

RESULT 11

FLGE TREPH
 ID FLGE TREPH STANDARD; PRT; 463 AA.
 AC Q56326.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Flagellar hook protein flgE.
 GN Name=flgE;
 OS Treponema phagedenis.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=162;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Kazan 5;
 RX MEDLINE=94266716; PubMed=8206841;
 RA Limberger R.J., Sliwinski L.L., Samsonoff W.A.;
 RT "Genetic and biochemical analysis of the flagellar hook of Treponema
 phagedenis.",
 RL J. Bacteriol. 176:3631-3637 (1994).
 CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

[illegible][illegible]

Db 230 MEFYKVDNTWKARLSMTDSTQLSDVSGTGQNTQLPGNTELEFGFTPDGKLIVYVSDGV 289
Qy 276 -----NGELKNMTAFTPTGS---ATKDLNAPQAPLVNGLPOFSANFVGAGIQPLTLDFGI 328
Db 290 DSMNSGKLNKAKVFRIPGNPAIONFDNLNGL-AGVNGITQFSSDF----- 334
Qy 329 KSQONWAGAPASAAAIGTDIGKLPSPMPITQSGNSTARNSSSTRYSQDGPQGLV 388
Db 335 -----TTKAVKQDGYTWGYLE 350
Qy 389 DVTITSEGLQKYSNQVDFYNIPLARTSEDGLRREGNNHYSATLDSGGPEFLPGT 448
Db 351 SFSIDNSGTTIGVPSNGVRQPLARVATAVFNPAGLDKAGDTMFSYNSNGEPNIGEAGV 410
Qy 449 SNYGLSVNQLETSNVDMRSREVMNMIIOQGFQMNKSVTTADTMLOKALELKR 502
Db 411 QGRGKINAGLLENSVLDSCQFTDMIVTQRFQANSRTITTSQMIQEVGLKLR 464

RESULT 15
Q8F2C6
ID Q8F2C6 PRELIMINARY; PRT; 464 AA.
AC Q8F2C6
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Flagellar hook protein.
GN Name=flag; OrderedLocusNames=LA2848;
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiroaceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
EX MEDLINE=22598143; PubMed=127112204;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-P., Jiang H.-O., Jia J., Tu Y.-F.,
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-F., Ma W.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RT "Unique physiological and pathogenic features of Leptospira
RT interrogans revealed by whole-genome sequencing."
RL Nature 422:888-893(2003).
DR EMBL; AE011450; AAN50047.1;
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; P:motor activity; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR001444; Flag_bb_rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF00460; Flg_bb_rod; 1.
DR PROSITE; PS00388; FLAGELLA_BB_ROD; UNKNOWN_1.
KW Complete proteome; Flagellum.
SQ SEQUENCE 464 AA; 50349 MW; 7EDD5A84709F84F CRC64;

Query Match 21.3%; Score 550; DB 2; Length 464;
Best Local Similarity 27.9%; Pred. No. 2.3e-25;
Matches 149; Conservative 84; Mismatches 199; Indels 102; Gaps 12;

Qy 1 MMSLFTGATGKMTSTGLTGVNNANANTIGYKQOVVFDLFSQDLAIGS-----TG 55
Db 1 MMSLYSGVGLKNNHQRMDVGNISNVNTHGFKTERVTFQDMISQELRGASEPKENIG 60
Qy 56 SQGNQAGMGAQVGSVRTIFTQAFEPGNSVTDLAIGGKGFQVTLDEKVVHTRAGNFRF 115
Db 61 GVNPOQVGLSLIAADIKMTQSLQTTGKNTDVAISGEGFIVKDGKQFYTRAGFNL 120
Qy 116 TDQG-FLNDPSGFTLMG--SRISNNPNKIKTELEIQLDFNDPTVAKSPAKTSTALNAV 172

Db 121 DKNGYVNPANGLKVGQWNSRLDDKGNKYINSAASIE-DIIIPVYSKEPARATSOIDPKS 179
Qy 173 NLGDS-----DKTQSEANPYFALLESKWNGTPISTSNYSYAQPMVYDQGNSHDIT 227
Db 180 NNSAPAPVPDQOEIT---AMIND-----PDKWRGHVTTINTFDDQGIQREFK 229
Qy 228 VYFDGAPSTGSKTFEYLVAMNPSDGAASGTSA-----GLMSGTMTPSS--- 275
Db 230 MEFYKVRDNTWKARLSMTDSTQLSDVSGTGQNTQLPGNTELEFGFTPDGKLIVYVSDGV 289
Qy 276 -----NGELKNMTAFTPTGS---ATKDLNAPQAPLVNGLPOFSANFVGAGIQPLTLDFGI 328
Db 290 DSMNSGKLNKAKVFRIPGNPAIONFDNLNGL-AGVNGITQFSSDF----- 334
Qy 329 KSQONWAGAPASAAAIGTDIGKLPSPMPITQSGNSTARNSSSTRYSQDGPQGLV 388
Db 335 -----TTKAVKQDGYTWGYLE 350
Qy 389 DVTITSEGLQKYSNQVDFYNIPLARTSEDGLRREGNNHYSATLDSGGPEFLPGT 448
Db 351 SFSIDNSGTTIGVPSNGVRQPLARVATAVFNPAGLDKAGDTMFSYNSNGEPNIGEAGV 410
Qy 449 SNYGLSVNQLETSNVDMRSREVMNMIIOQGFQMNKSVTTADTMLOKALELKR 502
Db 411 QGRGKINAGLLENSVLDSCQFTDMIVTQRFQANSRTITTSQMIQEVGLKLR 464

RESULT 16
AAS69795
ID AAS69795 PRELIMINARY; PRT; 464 AA.
AC AAS69795;
DT 31-MAR-2004 (TREMBlrel. 27, Created)
DT 31-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 11-MAY-2004 (TREMBlrel. 27, Last annotation update)
DE Flge.
GN FLGE OR LIC11188.
OS Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar
OS Copenhagen).
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiroaceae; Leptospira.
OX NCBI_TaxID=44275;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fiocruz L1-130;
EX PubMed=15028702;
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,
RA Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A.,
RA Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrier H.,
RA Coutinho L.L., Degraeve W.M., Dellagostin O.A., El-Dorry H.,
RA Ferro E.S., Ferro M.I.T., Furlan L.R., Gamberini M., Gigliotti E.A.,
RA Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,
RA Jeronimo S.M.B., Junqueira-de-Azevedo I.L.M., Kimura E.T.,
RA Kuranae E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L., Nunes L.R.,
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schrieffer A.,
RA Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,
RA Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Sluys M.A.;
RT "Comparative genomics of two Leptospira interrogans serovars reveals
RT novel insights into physiology and pathogenesis."
RL J. Bacteriol. 186:2164-2172(2004).
DR EMBL; AE017291; AAS69795.1; -
SQ SEQUENCE 464 AA; 50349 MW; 7EDD5A84709F84F CRC64;

Query Match 21.3%; Score 550; DB 2; Length 464;
Best Local Similarity 27.9%; Pred. No. 2.3e-25;
Matches 149; Conservative 84; Mismatches 199; Indels 102; Gaps 12;

Qy 1 MMSLFTGATGKMTSTGLTGVNNANANTIGYKQOVVFDLFSQDLAIGS-----TG 55
Db 1 MMSLYSGVGLKNNHQRMDVGNISNVNTHGFKTERVTFQDMISQELRGASEPKENIG 60
Qy 56 SQGNQAGMGAQVGSVRTIFTQAFEPGNSVTDLAIGGKGFQVTLDEKVVHTRAGNFRF 115
Db 61 GVNPOQVGLSLIAADIKMTQSLQTTGKNTDVAISGEGFIVKDGKQFYTRAGFNL 120

GN	Name=FlgE3; OrderedLocusNames=TTE1433;
OS	Thermoanaerobacter tengcongensis.
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC	Thermoanaerobacteriaceae; Thermoanaerobacter.
NCBI	TaxID=119072;
[1]	_TaxID=119072;
RN	SEQUENCE FROM N.A.
RP	STRAIN-MB4;
RC	MEDLINE=121992816; PubMed=11997336;
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;
RT	"A complete sequence of the T. tengcongensis genome.";
RL	Genome Res. 12:689-700(2002).
DR	EMBL; AE013102; AAM24655.1;
DR	GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.
DR	GO; GO:0003774; P:motor activity; IEA.
DR	GO; GO:0005198; P:structural molecule activity; IEA.
DR	GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR	InterPro; IPR010930; DUF1078.
DR	InterPro; IPR011491; Flae.
DR	Pfam; PF06429; DUF1078; 1.
DR	Pfam; PF07559; Flae; 1.
DR	Pfam; PF00460; Flg_bb rod; 1.
DR	PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
KW	Complete proteome; Flagellum.
SQ	SEQUENCE 415 AA; 43973 MW; 9EC1653E38510ECF0 CRC64;
Query Match	20.4%; Score 528.5; DB 2; Length 415;
Best Local Similarity	28.9%; Pred. No. 4e-24;
Matches 148; Conservative 77; Mismatches 180; Indels 107; Gaps 12;	
QY	1 MWGLSIFGATGMKTHSTGLGTVSNNTANANTICYKQQVVFODLSQDL-----AIGSTG 55
DB	1 MLRSMYSAISQLQHARLDVGNNIANVTNGVKASMTFKETFSQIKGASAPQGNGG 60
QY	56 SQGENQAQMGAGVGSVRTIFTQGAFFPGNSVTDLAIGCKGFQVLTLEDKVHYTRAGNRFR 115
DB	61 GTNPQQICGLGVAIASIDTLFTRGCAORTNDPTLSDGDNGFFIVSGESILYTRAGNESF 120
QY	116 TDGFLNDPSGFTLMGSRISNNPNKKET--LEPIOLD--FN-DPTVAKSPAKTSTALNA 170
DB	121 DSNGDLVTPGYKVLGHMSTDGKTVDTCNLVPISLKNWFGEFS-----TTTQLEI 173
QY	171 VVNIGDSTDTQTSEANPYFALLESWKNGTPTPISTSNYSYAQPMRVYPQQGNSHDITVVF 230
DB	174 GGNNLNASTGICQS-----ISY--NIITYPTQGGSHVATITF 207
QY	231 DGAFSSGTGKTFEYLVMNPSEDGSAASCTDSAGLLMSGTMTFSSNGELKNKTAFTPTGS 290
DB	208 TRQDPNTWN-----WSVSSQDPYISSVSAG--STATITFGADGKIAPT--QATGT 254
QY	291 ATKOLNAWQAPLVNGLPOFSANFVGAGIQPLITLDFGIKSQNMMWAGAPASAAAITGDI 350
DB	255 LTFNNMT-----AVTNQAIGPVNIDL 276
QY	351 KLPMSMPIOTSSGNSTARNGSSSTRYSODGYPQGDLDVDTITSBKLGKYKGSYQVVDF 410
DB	277 KL-----TMFSTETDLRELKNGNEAGSLQINIDKYGVVVSGIYSGNRQVI 323
QY	411 YNIPLARFTSEDLRRGNHNSYATLDSGGPEFLPGTINSYKULSVNOLETSNVDMGREM 470
DB	324 GQIALADFQPMGLEKVGKTMFTINTVNSEPMIGAAGSGTRGSINPGTLEMSNVLNAEF 383
QY	471 VNMIIRGQFMNKSQVTTADTMLQALEKR 502
DB	384 VDMITTQRGFQANAKVITVSDEMLQDLVNMR 415
RESULT 20	
Q914P9	PRELIMINARY; prt; 462 aa.
ID Q914P9	

Q914P9;
01-MAR-2001 (TrEMBLrel. 16, Created)
01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Flagellar hook protein FlgE.
GN Name=flgE; OrderedLocusNames=PA1080;
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
[1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzer L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen";
RL Nature 406:959-964(2000).
DR EMBL; AE004539; AAC04469.1; -.
DR PIR; F83510; F83510.
DR GO; GO:000288; C:flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR001444; Flag_bb rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF00460; Flg_bb rod; 1.
KW Complete proteome; Flagellum.
SQ SEQUENCE 462 AA; 48336 MW; 519616DABACE9C48 CRC64;

Query Match 20.4%; Score 527.5; DB 2; Length 462;
Best Local Similarity 28.5%; Pred. No. 5.4e-24;
Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;

QY 4 SLFIGATGKHTSTGLGTGVTNNIANANTIGYKQOVVFQDLFSQDLAIGSTGSGPQAG 63
DB 2 SFNIGLSGQAASSGLNVTGNNIAGTVGPKRAEFADYAAASV-LGS-----GSPQ 56

QY 64 MGAQVGSVRTIPTQGAFFPGNSVTDLAIGKGFQVLTEDKVHYTRAGNPRFTODGFLND 123
DB 57 SGVLLSDVSQMPKQGNIDSTNSVLDLAINGNGFFVTSNNGAISTRAGYFNTDKQFIVD 116

QY 124 PSGFILMSRISNNPIKKEFILEPIQLD-FNDPTVAKSPAKTALNAVNLGSDTDKQ 182
DB 117 NNGYRLOGYAVGPNQQLQGVVTDLKVRAQAPQATSSIQSYNLNSTLK-----P 168

QY 183 SEANPYFALLSWKNGTPPTSTNYSVAQPMRVYDQGNSHDITVVF---DCGAPST-- 237
DB 169 PTVTFP-----DPSDATYNSSSGLGYDSQGNHNTMSQFFKNEPDPNATPP 216

QY 238 ---GSKTEYILV-AMNPSEDSAAAGTSDSAGLLSGMTFTSSNGELKNMTAFTPTGS--- 290
DB 217 IPENSWTKVLIDGVNPLDPSNKT-----MSFNVTDPASQMTSVRA---PDGSGT 266

QY 291 -----ATKDLNAPQPLVNLGFPQSANFVAGIQPLTDLFGKQSNWAGAPASAAA 344
DB 267 PGFIDATNVIQSPS---TGNPPTP-----GTGWIPAASD-GKTPPTPAWNGATGAAG 318

QY 345 IGTDIGKLPSMMP:QTSGNSTARGSSSTRYS-----QDGYPGDLVDVITTS 395
DB 319 ISFDWRK-----TTQSTAFQAQSNPIQDGYTTCQLAGLEIDDT 356

QY 396 GKLAGKYSNVOVDYFNIPARFTSEGLRREGNNHYSATLDSGGPFGPLGTSNKGKLS 455
DB 357 GVIFARYINGKVGQGVVLANFANIQGLTPIGKTSWVQSSGEPVAGAPRSGTILGALQ 416

QY 456 VNOLETSNVDSREVMNMIIRQGFOMNSKSVTTADTMLOKALELK 501
DB 417 SGALEASNVDISNELVNLIVHQRNYQANAKTIQEDAVTQTIINLR 462

RESULT 21
FLGE_BORBU
ID FLGE_BORBU STANDARD; PRT; 442 AA.
AC Q44767; Q44734; Q44899;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Flagellar hook protein flgE.
GN Name=flgE; OrderedLocusNames=BB0283;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA Dunn J.J., Butler-Loffredo L., Kieleczawa J., Medalle J., Luft B.J.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=HB19;
RA Old I.G.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
Richardson D.L., Peterson J.D., Keflavage A.R., Quackenbush J.,
Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
Gocayne J.D., Weidman J.F., Utterback T.R., Wattley L., McDonald L.A.,
Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi";
RL Nature 390:580-586(1997).
[4]
RP SEQUENCE OF 18-207 FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA Limberger R.J., Slivenski L.L.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins
family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U43739; AAA85606.1; -;
DR EMBL; U43849; AAA87351.1; -;
DR EMBL; AB001137; AAC66665.1; -;
DR EMBL; U19712; AAA61738.1; -;
DR PIR; C70135; C70135.
DR TIGR; BB0283; -;
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR001444; Flag_bb rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF00460; Flg_bb rod; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
KW Complete proteome; Flagellum.
FT VARIANT 18 18 R -> I (in strain HB19).
FT VARIANT 119 119 D -> Y (in strain HB19).

FT VARIANT 174 174 V -> I (in strain HB19).
FT VARIANT 192 192 N -> S (in strain HB19).
SQ SEQUENCE 442 AA; 47389 MW; 561AC092B72C69BE CRC64;
Query Match 19.7%; Score 509; DB 1; Length 442;
Best Local Similarity 28.2%; Pred. No. 6.7e-23;
Matches 150; Conservative 71; Mismatches 190; Indels 120; Gaps 13;
QY 1 MMSLSFGATGKTHSTGLTGVSNINANTIGYQQQVFDLPFSDLAIGS-----TG 55
Db 1 MMSLSYGVSGLQNHQTHQMDVGVNNIANVTIGFKKGVNFQDMISQISGASRPTDARG 60
QY 56 SQCPNQAAGCAQVGVRTFTQGAPEFGNSVTDLAIGGKGFQVTLDEKHYTRAGNPRF 115
Db 61 GTNPKQVGLGMNVASIDTHTTQGAQSTQKASDLGVSGNGFFLLKEGKNLYTRAGAFDV 120
QY 116 TQDGLNDSGFTLMGSRISN--NPNIKKETLEPIQLDND---PTVAKSPAKTSTALNA 170
Db 121 DSDRLHVNPN-----GMRIQGMWARDLEGEKVINTASDIEDLIPIGDKGAKTKNVT 176
QY 171 VVNLGSDTKTQSEANPYFALLESWKNGCTPPISTSNYSYAQPMRVYDQGNSHDITVYF 230
Db 177 ACNLDKRLPLIQGANPADIAIGTGWVVKNS-----LYDSFGN---VSVL- 217
QY 231 DGAPSGTGKTEYLVAMNPSEDGSAASGTDAGLLMSGTWFSSNGELKNWTAFTPTGS 290
Db 218 -----ELR----- 220
QY 291 ATKDLNA---WOPAPLVNGLPOFSANFVGAGIQPLTLDP-----GIKSOQNMWAG 337
Db 221 VVKDLATPNLWATVLNG--EQNSNF-----TLGPDNEGALASLNGPGQKGDILQ 270
QY 338 APASAAAIGTDIGKLSVMPIQTSQ-----NSTARNSSSTRYSDQGYPOGDLVDV 390
Db 271 IPITFNVLGANVEGEQQTQVNLKLTGVSTYDTSITQFADSSSTKAIITQDGYGMGYENY 330
QY 391 TITSEKLGKYSNSQVDFYNIPLARTSEDLREGNNHYSATLDSGGPEFGFLPGTSN 450
Db 331 EIDQNGVIYVSGIRRLGXIALASFWNPGLAKSGDTNFEVTSNGSVQVRIGETGLAG 390
QY 451 YGKLSVNOLETSNVDMRSREWNMIILQRFQWNSKSVTTADTMLOKALELK 501
Db 391 LGSIRSGVLEMANVDLAEOTDMIVTQRFQANAKTITTSOLLQELVRLK 441
RESULT 22
Q8XSX2 PRELIMINARY; PRT; 401 AA.
AC Q8XSX2
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE PROBABLE FLAGELLAR HOOK PROTEIN FLGE.
GN Name=flge; Synonyms=RS00745; OrderedLocusNames=RS0345;
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823952;
RA Saranoubat M., Genin S., Ariguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billaut A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
BL Nature 415:497-502(2002).
DR EMBL; AL646078; CAD17496.1; --
DR GO; GO:0009288; C:Flagellum (sensu Bacteria); IEA.

DR GO; GO:0003774; E:motor activity; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; FlaeE.
DR InterPro; IPR001444; Flag_bb_rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; FlaeE; 1.
DR Pfam; PF00460; Fig_bb_rod; 1.
DR PROSITE; PS00888; FLAGELLA_BB_ROD; 1.
KW Complete proteome; Flagellum; Plasmid.
SQ SEQUENCE 401 AA; 41576 MW; 6F18569259DACC0 CRC64;
Query Match 19.4%; Score 502; DB 2; Length 401;
Best Local Similarity 27.4%; Pred. No. 1.6e-22;
Matches 137; Conservative 85; Mismatches 168; Indels 110; Gaps 15;
QY 8 GATGKTHSTGLTGVSNINANTIGYQQQVFDLPFSDLAIGS-----DQIGQGV 67
Db 6 GLSGDLAASKNLDVIGSVNANVTGKYSTAEFGDYARSL-VGAS-----DQIGQGV 60
QY 68 VGSVRTFTQGAPEFGNSVTDLAIGGKGFQV--TLEDKVHYTRAGNFRFTQDGLNDPS 125
Db 61 VTKVSQSTQGVNVTGPNPLDIAINGTGFRVMDASSQVSYTENGQFQTDKNGYIISAT 120
QY 126 GTFLMGSRISNPNKIKETLEPIQLDNDPTVAKSP-AKTSTALNAVNVNLDGSDTKTQSE 184
Db 121 GQNLTYGVGDATGKINTAVLTNLQIPVND-----LAPLATNTAFS---INL-DAAGTVPT- 172
QY 185 ANPYFALLESWKNGCTPPISTSN---YSYAQPMRVYDQGNSHDITVYFGAPSGSTGSKT 241
Db 173 -----TTFSSATNSATFNHVSVEQYDGTGTSMLTNYV--VRTAAGWDV 215
QY 242 FYLVAMNPSEDGSAASGTDAGLLMSGTWFSSNGELKNWTAFTPTGSATKDLNAPQA 301
Db 216 YSQVDPGANPTGPNVTS-----LTFNSSGVL-----TSSPSKVAFAF--- 252
QY 302 PLVNGLPQFSANFVGAGIQPLTLDFGIKSQQNMWAGAPASAAICTDIGKLSMMPPIQTS 361
Db 253 -----AGNSIANMDF-----TGT----- 265
QY 362 SGNSTARNGSSSTRYSQDGYPOGDLVDVTTITSEKLGKYSNSQVDFYNIPLARTSE 421
Db 266 -----TQYGGGFNDTTSQDGYATGRLASYSVGTGTTTGRYSNGRTSLTGLGIANTNFKAP 321
QY 422 DGLREGNNHYSATLDSGGPEFGFLPGTSGYKLSVNOLETSNVDMRSREWNMIILQRFQ 481
Db 322 DGLQNGVQWVETAEFGAPQWGPFGMSFGLOSSAVEQSNVLSAELVNMIVAQRSYQ 381
QY 482 MNSKSVTTADTMLOKALELK 501
Db 382 ANAQTKTETDITLQTLVSMR 401
RESULT 23
Q8FIQ4 PRELIMINARY; PRT; 401 AA.
AC Q8FIQ4
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Flagellar hook protein flge.
GN Name=flge; OrderedLocusNames=cl345;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasco D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

Query Match 19.3%; Score 498; DB 1; Length 402;
Best Local Similarity 28.7%; Pred. No. 2.8e-22;
Matches 144; Conservative 76; Mismatches 167; Indels 114; Gaps 14;

QY 10 TGMKTHSTGLTVSNINANTIGYKQOVVQDLFSQDLAIGSTGSGPQAGVQV 69
DB 7 SGLNAAATMLDVIGNNIANSATYGFSGTASADMF 54
QY 70 SVRTITFOGAFPGNSVTDLAIGGKGFQ-VTLEDKVHYTRAGNFRFTQDGFNDPSGFT 128
DB 55 GITQDFTDGTNTGRLDVAISQNGFFRLVDSNGSVFYSRNGQFKLDENRNLVNMQMG 114
QY 129 LMGSRISNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTTOSEANPY 188
DB 115 LTGYPATGTPPTIOQGANPAPITPNTLMA---AKSTTTASMQINL-NSTDVPSKT--- 167
QY 189 FALLESWKNGTTPPIST---NYSVAQPMRVYDOQGNSHDITVYFDGAPSGTSGKTFEYL 245
DB 168 -----PFSVDADSYNKKGTVTYVDSQGNADHNMVYF-----VKTQDNEWA 208
QY 246 VAMNPSEDSAGSAGTDSAGLLSGMTFTSSNGELK-----NMTAFTPTGSAKDLNANQP 300
DB 209 VYTHDSDDPAAAPTAA-----STLKFENGLISGGTVNITTTGTINGATA----- 255
QY 301 APLVNLGPOFANFVGAGIQPLTDFGIKSOQNMWAGAPASAAAIGTIDIGKLPSPMPTQ 360
DB 256 -----ATFSLSFLNS-----MQQN-----T 270
QY 361 SSGNSTARNSSSTRYSODGYPOGDLVDVITTSSEKLGKYSNSQVVDVFNIPARFTS 420
DB 271 GANNIVATN-----QNGYKPGDLVSYQINNDGTVVGYSNEQEOVLQIVLANFAN 321
QY 421 EPLRREGNHYSATLDSGGPFGPLGTSNYGKLSVNOLETSNVDMSPRYNMIIIOQGF 480
DB 322 NEGLASQGNVWAATQASGVALLGTAGSGNFGKLTNGALEASNDLSKELVNMIVAQENY 381
QY 481 QMNSKSVTTADTLMQKALELK 501
DB 382 QSNAGTIKTQDQILNTLVNLR 402

RESULT 25
Q7AF93 PRELIMINARY; PRT; 401 AA.
ID Q7AF93
AC Q7AF93;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Flagellar hook protein flgs.
GN OrderedLocusNames=ECs1454;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Tida T., Takami H., Honda T., Sakakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AP002555; BAB34877.1; --
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF00460; Flg_bb rod; 1.
DR PROSITE; PS00588; FLAGELLIN_BB_ROD; 1.

KW Flagellum.
SQ SEQUENCE 401 AA; 42014 MW; 453B39179717CB14 CRC64;
Query Match 19.1%; Score 493; DB 2; Length 401;
Best Local Similarity 28.7%; Pred. No. 5.6e-22;
Matches 142; Conservative 76; Mismatches 174; Indels 102; Gaps 15;

QY 10 TGMKTHSTGLTVSNINANTIGYKQOVVQDLFSQDLAIGSTGSGPQAGVQV 69
DB 8 SGLNAAATMLDVIGNNIANSATYGFSGTASADMF 55
QY 70 SVRTITFOGAFPGNSVTDLAIGGKGFQ-VTLEDKVHYTRAGNFRFTQDGFNDPSGFT 128
DB 56 GITQDFTDGTNTGRLDVAISQNGFFRLVDSNGSVFYSRNGQFKLDENRNLVNMQGLQ 115
QY 129 LMGSRISNNPNKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKTTOSEANPY 188
DB 116 LTGYPATGTPPTIOQGANPAPITPNTLMA---AKSTTTASMQINL-NSTDVPSV--- 167
QY 189 FALLESWKNGTTPPIST---NYSVAQPMRVYDOQGNSHDITVYFDGAPSGTSGKTFEYL 248
DB 168 -----NADSNADSYNKKGSVTVFDSQGNADHNMVYF-----VKTGDNHNVY--- 211
QY 249 NPSEDSAGSAGTDSAGLLSGMTFTSSNGEL-KWMTAFTPTGSAKDLNANQPAPLVNGL 307
DB 212 --TQSDSDPTGTABPAM---KLVENANGVLTSNPTENITGA---INGAEP--- 255
QY 308 POFSAFVGAGIQPLTDFGIKSOQNMWAGAPASAAAIGTIDIGKLPSPMPTQSSGNSTA 367
DB 256 -TFSLSFLNS-----MQQN-----TGANNIVA 276
QY 368 RNSGSTRYSODGYPOGDLVDVITTSSEKLGKYSNSQVVDVFNIPARFTSBDGLRRE 427
DB 277 T-----TQNGYKPGDLVSYQINNDGTVVGYSNEQEQQLQIVLANFANNEGLASE 327
QY 428 GNNHYSATLDSGGPFGPLGTSNYGKLSVNOLETSNVDMSPRYNMIIIOQGF 487
DB 328 GDNVWSATQSSGVALLGTAGSGNFGKLTNGALEASNDLSKELVNMIVAQENYQSNQTI 387
QY 488 TTADTLMQKALELK 501
DB 388 KTQDQILNTLVNLR 401

RESULT 26
Q8X8L1 PRELIMINARY; PRT; 401 AA.
ID Q8X8L1
AC Q8X8L1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 26, Last sequence update)
DE Flagellar biosynthesis, hook protein.
GN Name=flgE; OrderedLocusNames=z1714;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11208551;
RA Ferna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
DR EMBL; AF005317; AAGS5822.1; --
DR PIR; B85670; B85670.
DR PIR; F90810; F90810.
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.


```
Db 275 VAT-----TQGYKPGDLSVYQINDGTGVVGNYSNQTQLLQGVILANFANNEGLA 325
Qy 426 REGNNHYSAILDSCGPFGLPRTSNYKLSVNOLETNSVDMSEVMNMIIOGFQWNSK 485
Db 326 SEGNNVWSATQSSGVALGTAGTGNFTLTNGALEASNDLSKELVNMVIAQRYOSNAQ 385
Qy 486 SVTTADTMLOKALELK 501
Db 386 TIKTQDQILNTLVNLR 401

RESULT 28
FLGE_HELPJ
ID _FLGE_HELPJ STANDARD; PRT; 718 AA.
AC Q92KYO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Flagellar hook protein flgB.
GN Name=flgB; OrderedLocusNames=JHP0804;
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682; DOI=10.1038/16495;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown B.D., Doig P.C.,
RA Smith D.B., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- SIMILARITY: Belongs to the flagella basal body rod proteins
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001509; AAD06365.1; -
DR PIR; G71888; G71888.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR010810; Flagellin IN.
DR InterPro; IPR001444; Flag_bb_rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07196; Flagellin IN; 1.
DR Pfam; PF00460; Flag_bb_rod; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; FALSE_NEG.
KW Complete proteome; Flagellum.
SQ SEQUENCE 718 AA; 76280 MW; DFGA0500D8EF481D CRC64;

Query Match 18.8%; Score 485; DB 1; Length 718;
Best Local Similarity 25.0%; Pred. No. 3.8e-21;
Matches 181; Conservative 94; Mismatches 22; Indels 226; Gaps 22;

Qy 1 MMGSLFIGATGKMTSTGLGTGVSNMIANANTIGYKQOVVFDLFSQDLAIGS-----TGS 56
Db 1 MRLSWGVNGVNGAHQIALDIENNNIANVNTGFKYSASFVDMLSQVKLATAPYKNGL 60

Qy 57 QGPN--QAGMGAVGSVRTITFGAFEPGNSVTDLAIGKGFQFVTLBKV--HYTRAGN 112
Db 61 AQONDPSVLGVGVDAATTKIFSGQNIQNTDVKTDLAIQDGGFFIISPRGITRFTDGE 120

Qy 113 PFPTQDGLNDPSGTLM-----GSRISNNNIKKETLEPIQLDFNDPTVAKSPAK 163
Db 121 FLFDGSGSLVTGGLVVGQWVRNGSDTGKNGSDTDALKVDNTGPLENTRIDPGV-NPAR 179
```

```
Qy 164 TST--ALNAVNLGDSITKTOSEANPYFALLESWK--GNGTPPISTSNYSYAQPMR----- 215
Db 180 ASNRISMEANLNRHADQTA--VFALDSSAKTPSDGINPVYDSGTNLQAQVAEDMGS 235
Qy 216 VTDOQN-----SHDITVFDGA-----PS-----STGSKTTEY 244
Db 236 LYNEGDGALLNENQGIWVYSKAKVVKDIIPSAENSTLENGVKISPTNDSAVRTSSL 295
Qy 245 LVAMNPSEDGSAAG-----TDSAGLLMGSTMTFFSSGELKN----- 281
Db 296 VAAKNAINAVKSGTGLEIAYLDOKQLRLNTNELDGEKIKNIIVTQAGTGAFANFLDGDK 355
Qy 282 -MTAF-----TPTGSATKDLNMQPAPLVNGLPQPSANFV-----GAGIQPLTL 324
Db 356 DVTAFKYSYTHSISPNADIGQPTTDLRALTIQHDANIVKDPSLADNYQDSAAISGVITIN 415
Qy 325 DFG-----IKSQNNW-----AGAPASAA 343
Db 416 QYGMFEINKONKNIKENLNFVSGYSDSVTNVLPNAMKGLNTASLISSGASASS 475
Qy 344 -----AIGT----- 347
Db 476 KETHATHATSIDVIDSLGTGHAMRIEFYRSGGAENFRVIVPEPGELVGGSAARPNFEG 535
Qy 348 -----DIGKPSMP-----IQTSSGNSTARGSSSTRYS-----O 379
Db 536 GRLHFNNDGSLAGMNPPLLPQDPKNGADAPQRIINLAFSGSGFDGLTSVDKISFYAIEQ 595
Qy 380 DGYVQGDLDVDTITSEGLQKYSNSQVVDYFNIPARFTSDEGLRREGNNHYSATLDSG 439
Db 596 NGYQAGDLMDVRFDSGVLGAFSGRTLALAAQVALANFANDAGLQALGNGVFSQTGNSG 655
Qy 440 GBEFLGPTGTSYNGKLSVNOLETNSVDMSEVMNMIIOGFQWNSKSVTTADTMLOKALE 499
Db 656 QALIGAANTGRRGSISGSKLESNDLSRLTNLIVQVQFOANSKAVTTSDQILNTLNLN 715

Qy 500 LKR 502
Db 716 LKQ 718

RESULT 29
Q92DV7
ID Q92DV7 PRELIMINARY; PRT; 411 AA.
AC Q92DV7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Lmc0697 protein.
GN OrderedLocusNames=lmo0697;
OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitouran A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstrek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voes H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591976; CAC98775.1; -
DR PIR; A11161; A11161.
```


OX NCBI_TaxID=1642;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLIP 11262 / Serovar 6a;

RX MEDLINE=2153729; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rueniok C., Amend A.,

RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,

RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duseurget O.,

RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,

RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,

RA Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordisk G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,

RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;

RT "Comparative genomics of *Listeria species*,"

RL Science 294:849-852(2001).

DR EMBL; AL596166; CAC95937.1; -.

DR ListerList; LIN0705; -.

DR InterPro; IPR010930; DUF1078.

DR InterPro; IPR011491; Flae.

DR InterPro; IPR001444; Flag_bb_rod.

DR Pfam; PF06429; DUF1078; 1.

DR Pfam; PF07559; Flae; 1.

DR Pfam; PF00460; Flg_bb_rod; 1.

DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.

KW Complete proteome.

SQ SEQUENCE 411 AA; 42836 MW; 3272D338BB4E1A7C CRC64;

Query Match 18.7%; Score 484.5; DB 2; Length 411;

Best Local Similarity 27.0%; Pred. No. 1.9e-21;

Matches 137; Conservative 77; Mismatches 188; Indels 105; Gaps 137;

Qy 1 MMGSLFTGATGMKTHSTGLGTVSNINANANTIGYKQOVFPQDLFSQD-----LAIGSTGS 56

Db 1 MNQMTYAISGMNAPFOQALSVTSSNNINANTGYKKQVWFNDLLYQNTMGSVAGGLYAG 60

Qy 57 QGPNQAGCAQGVSRIFTOGAFPGNSVTDLAIGKGFFQV--TLDDKVHYTRAGNFR 114

Db 61 TNPMSFGSGKIGAILTDYTAGSPSTGTGRNKDAALQGRGFFIAGDNAGCNVIYTRDGSFA 120

Qy 115 FTQDGLNDPSGFTMGSRISNNPNIKKETLEPQLDFNDFTVAKSPAKTSTALNAVNL 174

Db 121 VSDNNYLTTQQKYVMGYATDKGNVLNGNLQPIQLPNSAIPGEATKNGSLGNIPLDW 180

Qy 175 GDSDTKTQSEANPFPALLESKNGKNGTPPISNTSYAQPVMVYQQQSHDITVYFDGA- 233

Db 181 GEK-DTTSSE-----LSVYNAGKHKLVNMKAAT 210

Qy 234 PSSTGSKTFEYLVAMPNPSGDSAASTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293

Db 211 PDASGNVSYEYEQM-----DGKALTPP-----VTGLTNYNAQGLTNPDA----- 251

Qy 294 DLNAWOPAPLVNGLPQFSANFVGAGIQPLTDFGIKQONWAGAPASAAAGTIDIGLKP 353

Db 252 -LKNIQINSTVNG-----KQVNM-----GLNLSGLT 276

Qy 354 SMMPITQSSGNSTARNGSSSTRYSQDGYPOGLDVLDTITSEGKLGKYSNVQVDFYNI 413

Db 277 NV-----GTNQVFSPTSDGKAATVKDVAVTDSGYAVSYSGDGTVPVAQL 322

Qy 414 PLARPTSEDLGRRCNNHYSATLDSGGGFGLPGTFSNYGKLSVNOLETNVDMSREVMNM 473

Db 323 AVATFSNEDGLVKNMGYETVPLSGSDAVYGVAGQAGGAGGSGSLEGSNDVLSREFVNL 382

Qy 474 IIIQRGFQMNKSVTTADTMLOKALEL 500

Db 383 MTYGSGFGCGNTKVIKRVADVMKQIVNL 409

RESULT 32
AAT03514

ID Q7UCX0 PRELIMINARY; PRT; 402 AA.
AC Q7UCX0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Flagellar hook protein FlgE.
GN Name=flgE; OrderedLocusNames=S1160;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
Kau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; A016981; AAP16590.1; -
DR GO; GO:0003288; C:flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR001444; Flag_bb_rod.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF00460; Flg_bb_rod; 1.
DR PROSITE; PS00568; FLAGELLA_BB_ROD; 1.
KW Flagellum.
SQ SEQUENCE 402 AA; 42051 MW; E13798711ED7B22F CRC64;

Query Match 18.7%; Score 483.5; DB 2; Length 402;
Best Local Similarity 28.4%; Pred. No. 2.1e-21;
Matches 141; Conservative 77; Mismatches 173; Indels 105; Gaps 15;
QY 10 TGMKTHSTGLTVSNINANTIGYKQQVVDLFSQDLAIGSTGSGPQAGGAGVQ 69
DB 8 SGLNAAATNLVDIGNANSATYGFKSTGSPADWFA-----GS-----KVGLGVKVA 55
QY 70 SVRTITQCAFEPGNSVTDLAIGKGFQ-VLEDKVHYTRAGNFRFTQDGLNDPSGFT 128
DB 56 GITQDFTDGTNTGRGLDVAISQNGFRLVDSNGSVFYSRNGQFKLDENRNLNTQGLQ 115
QY 129 LMGSRISSNNPNTKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSPTKTOSEANPY 188
DB 116 LFGYPTGTPPTIQOAGNPTNISITLMA---AKTTTASMQINL-NSSDPLPT-----166
QY 189 FALLESWKNGTPTISTGN---YSYAPQMRVVDQGNSHDITVYFDGAPSGSTKTEYL 245
DB 167 -----VTPESASNADSYNKKGSVTVFDSQGNADMSVYF-----VKTGDNWQ--209
QY 246 VAMNPSEDSASGSDTSAGLLMSGTMTFSSNGELKNMTAFTPTGSAKDLNANQAPLNV 305
DB 210 VYTQSSDPSNTAKT-----ATTLEFNAGTL--VDGAMANNIATGAINGAEPA----256
QY 306 GLPQSFANFVGAGIQLTDFGKISQQNWAGAPASAAAIGTDIGKLPMMPIQTSSGNS 365
DB 257 ---TPSLGFLNS-----MQQN-----TGANNI 275
QY 366 TARNSSSTRYSQDYPQGDVVDVITSEKLOGKYSNSQVDFYNIPLARFSEGLR 425
DB 276 VAT-----TQNGYKPGDLVSVQINDGTVVGNNSNETQLLQGLVLANFANNEGLA 326
QY 426 REGNNHYATLDSGPEFLGPTSNYKGLSVNQLSTNSVNDMSREVMNIIQRGFQNSK 485
DB 327 SEGDNVWSATQSSGVALLGTAAGTGNFGTLTNGALEASNVLSKELVNNIQAQRNYSNAQ 386

QY 486 SVTTADTMLOKALELK 501
DB 387 TIKTQOILNTRVNL 402
RESULT 34
Q83RT3
ID Q83RT3 PRELIMINARY; PRT; 642 AA.
AC Q83RT3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Flagellar biosynthesis, hook protein.
GN Name=flgB; OrderedLocusNames=SF1080;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,
Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015135; AAN42702.1; -
DR GO; GO:0002888; C:flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR001444; Flag_bb_rod.
DR InterPro; IPR005648; FlgB.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF00460; Flg_bb_rod; 1.
DR PROSITE; PS00568; FLAGELLA_BB_ROD; 1.
KW Complete proteome; Flagellum.
SQ SEQUENCE 642 AA; 66632 MW; 64278E020767D34 CRC64;
Query Match 18.7%; Score 483.5; DB 2; Length 642;
Best Local Similarity 28.4%; Pred. No. 4e-21;
Matches 141; Conservative 77; Mismatches 173; Indels 105; Gaps 15;
QY 10 TGMKTHSTGLTVSNINANTIGYKQQVVDLFSQDLAIGSTGSGPQAGGAGVQ 69
DB 248 SGLNAAATNLVDIGNANSATYGFKSTGSPADWFA-----GS-----KVGLGVKVA 295
QY 70 SVRTITQCAFEPGNSVTDLAIGKGFQ-VLEDKVHYTRAGNFRFTQDGLNDPSGFT 128
DB 296 GITQDFTDGTNTGRGLDVAISQNGFRLVDSNGSVFYSRNGQFKLDENRNLNTQGLQ 355
QY 129 LMGSRISSNNPNTKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGDSPTKTOSEANPY 188
DB 356 LFGYPTGTPPTIQOAGNPTNISITLMA---AKTTTASMQINL-NSSDPLPT-----406
QY 189 FALLESWKNGTPTISTGN---YSYAPQMRVVDQGNSHDITVYFDGAPSGSTKTEYL 245
DB 407 -----VTPESASNADSYNKKGSVTVFDSQGNADMSVYF-----VKTGDNWQ--449
QY 246 VAMNPSEDSASGSDTSAGLLMSGTMTFSSNGELKNMTAFTPTGSAKDLNANQAPLNV 305
DB 450 VYTQSSDPSNTAKT-----ATTLEFNAGTL--VDGAMANNIATGAINGAEPA----496
QY 306 GLPQSFANFVGAGIQLTDFGKISQQNWAGAPASAAAIGTDIGKLPMMPIQTSSGNS 365

```
Db 497 ---TFSLFLNS-----MQQN-----TCANNI 515
Qy 366 TARGSSSTRYSGDGPQGLVDVTTITSEKLGKQKYSNQVDFYNIPIARTSDEGLR 425
Db 516 VAT-----TONGYKPGDLVSYQINDDGTVYGNNSQEQTLGLQVILANFANNEGLA 566
Qy 426 REGNNHYSATLDGSGPFGLPGTSNYGKLSVNOLETSNVDMSREVMNMIIQRFQWNSK 485
Db 567 SEGNNVWSATQSSVALLGTAGTCNFGTLTNGALEASNVLDLSKELVNMIVAQRYNSNAQ 626
Qy 486 SVTTADTMLOKALELK 501
Db 627 TIKTQDQILNTRVNL 642

RESULT 35
Q6D6H6 PRELIMINARY; PRT; 405 AA.
ID Q6D6H6
AC Q6D6H6
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Flagellar hook protein.
GN Name=flgE; Synonyms=flak; ORFNames=ECAL705;
OS Erwinia carotovora subsp. atroseptica SCRI1043.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=218491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI1043;
RA Bell K.S., Sebahia M., Pritchard L., Holden M., Hyman L.J.,
RA Holvea M.C., Thomson N.R., Bentley S.D., Churcher C., Mungall K.,
RA Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,
RA Praser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,
RA Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,
RA Salmund G.P.C., Birch P.R.J., Barrall S.G., Parkhill J., Toth I.K.,
RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; BX950851; CAG74611.1;
KW Flagellum.
SQ SEQUENCE 405 AA; 41910 MW; 687FB7C32380814B CRC64;

Query Match 18.6%; Score 482; DB 2; Length 405;
Best Local Similarity 28.6%; Pred. No. 2.6e-21;
Matches 142; Conservative 75; Mismatches 176; Indels 104; Gaps 16;

Qy 10 TGMKTHSTGLTVSNINANTIGYKQQVVFQDLFSODLAIGSTGSGPNQAGMGAQVG 69
Db 8 SGLRAASNNLDVIGNNIANSATVGFKSGNVTFADWEA-----GS-----KVGWGVKVA 55
Qy 70 SVRTITFOGAFEPGNSVTDLAIGKGFQVLTED---KVHYTRAGNFRFTQDGLNDPSG 126
Db 56 SVLQDFGNGTVTSSSRDLDAISGGGFYR--LQDTNGSTYYSRNGQFPLMGRNIVN-AQG 112
Qy 127 FTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVMLGSDTDKTOGEAN 186
Db 113 MOLTGYFVAGTPPTVQTGADPVLTPVPGDMLAS--QTSVA-SIKANL-KSSDSVPINA- 167
Qy 187 FYFALLESWKNGTTPPSTISNYSYAQPMRVYDQGNSHDITVYFDGAPSPSTGSKTFEVLV 246
Db 168 -----W-----ATTPGAEGTVNSKALTITVDSQGNVHNFYFKYKANTNW-QTYAKDD 215
Qy 247 AMNPEDCSAASGDSAGLLMGTTTFSSNGELKN--MTAETPGSATKDLNAQAPL 304
Db 216 SINPATYONA-----GTLPFANGALSTGTAHTP----- 246
Qy 305 NGLPOFSANFVAGIQPLTLDPGIKSQNNWAGAPASAAAIGTDIGKLPWMPITQSSGN 364
Db 247 -----FTLNLG-----TNGAANGT-----FTLNLG 268
Qy 365 STARNSSSTRYSQDGPQGLVDVTTITSEKLGKYSNQVDFYNIPIARTSDEGL 424
Db 269 SVQONTYSYKSTQNGFAPGSLTGFAINDDGTIEGYSNGOKQALQILLASFANPEGL 328
```

```
Qy 425 REGNNHYSATLDGSGPFGLPGTSNYGKLSVNOLETSNVDMSREVMNMIIQRFQWNS 484
Db 329 SPEGDNANSETASSQAVVGLAGTSLGKLGKSTESSNVLDLSKELVNMIVAQRYNSNA 388
Qy 485 KSVTTADTMLOKALELK 501
Db 389 QTIKTQDSILOTLVSLR 405

RESULT 36
Q7M8B3 PRELIMINARY; PRT; 716 AA.
ID Q7M8B3
AC Q7M8B3
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE FLAGELLAR HOOK PROTEIN.
GN Name=flgE; OrderedLocusNames=WS1758;
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSMZ 1740;
RX MEDLINE=22882897; PubMed=14500908;
RA Baar C., Eppinger M., Raddatz G., Simon J., Lanz C., Klimmek O.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
DR EMBL; BX571661; CAB10780.1;
DR GO; GO:0003288; C:flagellum (sensu Bacteria); IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0001598; F:structural molecule activity; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR InterPro; IPR010930; DUF1078.
DR InterPro; IPR011491; Flae.
DR InterPro; IPR010810; Flagellin_IN.
DR Pfam; PF06429; DUF1078; 1.
DR Pfam; PF07559; Flae; 1.
DR Pfam; PF07196; Flagellin_IN; 1.
DR Pfam; PF00460; Flg_bb_rnd; 1.
DR Complete Proteome; Flagellum.
SQ SEQUENCE 716 AA; 74684 MW; 28C478F91D30C605 CRC64;

Query Match 18.6%; Score 482; DB 2; Length 716;
Best Local Similarity 24.8%; Pred. No. 5.8e-21;
Matches 183; Conservative 79; Mismatches 217; Indels 260; Gaps 21;

Qy 1 MMSLFIATGKTHSTGLTVSNINANTIGYKQQVVFQDLFSODLA-----GSTG 55
Db 1 MLRLMSGVSGMQAHOQVALDVEGNINANTYNTTGFKYSRANFSDMLSQVNRATSPYGGLG 60
Qy 56 SQGPNQAGMGAQVGSVRTITFOGAFEPGNSVTDLAIGKGFQVLTED---KVHYTRAGNF 113
Db 61 GQNDYSIGLGTSLNSTTKIFGQSGIQDTTKMDLAIGDGGFFIVSGNGRTNAYTRDGAF 120
Qy 114 RFTQDGLNDPSGFTLMG-SRISNNPNIK-----KETLEPIQLDFNDPTVAKSPAK 163
Db 121 GFDAAGNMVNNAGYIVQGWTRDLNLSGSTSCYSDALYNVVDTTVPISGKIEPKQW-IPAK 179
Qy 164 TSTALNAVNL--GDSTDK-----TQSEANPYFALLSWKNGKNGTTPPSTISNYSYAQ 212
Db 180 ATTQVNLDAVLTAGDTITDKLGCMYALDTSVTAADGIAARYDSAGN-----KIQMAE 231
Qy 213 PMRV-YDQGN-----SHDITV----- 228
Db 232 DMGVLNFASGNALKSEGGQGVVSYQATATQAVATTGTTITLNGTITFTINDSTISGV 291
Qy 229 -----YFDGAPSTGSKTF-----EYLVAWNPSEDGSA-----ASG----- 259
```

Db 292 SSLVAQAQNAKSGTGEATGQGLRLVNDNSLDGNASKNVIIITASGALANFTAA 351
 QY 260 -----TDSAGLLMSGTMT 272
 Db 352 DNSITAFRAYTASDADSTSLRFTEDLRLALLODANNIKHGGYVDSGTNASVKVT 411
 QY 273 FSNGBELK-----NMATFTGSAATKDLNA-----297
 Db 412 INKTGMFEILNODGDTTITNLNLSLTVSSYDNTNVSFLFKSAMKGLNTGILVEGSSST 471
 QY 298 -----W-----OPALVNLGPQFSANF 314
 Db 472 SASLMAAKHTATTDIVDSLGNKHTLTVPRKVGPOKESFLVHPATFVNGSGRPNYF 531
 QY 315 VG-----AGIQPTLDFGKSNQNMWAGAPASAAAIGTDIGKLPNMPIQTSSG 363
 Db 532 EGRVTFGSDGGLTGMNPTIQFNPKS-----GA-SSPORIDLDG-----VSGTFQ 577
 QY 364 NSTARNGSSSTRYSODGYPQGLVDVITTSSEKLGKYSRQVYDFYNIPLABETSEDG 423
 Db 578 GUTSDKXGATGNIYONGYQSGVLEMDRPFDSNGVLIGFSGKGLALAOVAIAGFTNNGG 637
 QY 424 LRREGNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMRSYMNMIIOQGFQMN 483
 Db 638 LQAGSNLFSQANSGEPPVGTAGSGGRKISPSALEMSNVDSLSRLTQLIIVQGFQAN 697
 QY 484 SKSVTTADTMLOKALELR 502
 Db 698 SKVTTSDDQILNTLLQLK 716

RESULT 37

Q8P9B6 PRELIMINARY; PRT; 407 AA.
 ID Q8P9B6
 AC Q8P9B6
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Flagellar biosynthesis, hook protein.
 GN Name=flagE; OrderedLocusNames=XCC1949;
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RA MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Forughieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RL Nature 417:459-463(2002).
 DR EMBL; AF012295; AAM41238.1;
 DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
 DR InterPro; IPR010930; DUF1078.

DR InterPro; IPR011491; FlaE.
 DR InterPro; IPR001444; Flag_bb_rod.
 DR Pfam; PF06429; DUF1078; 1.
 DR Pfam; PF07559; FlaB; 1.
 DR Pfam; PF00460; Flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_rod; 1.
 KW Complete proteome; Flagellum.
 SQ SEQUENCE 407 AA; 42710 MW; 307EB1287BE2B079 CRC64;
 Query Match 18.6%; Score 481; DB 2; Length 407;
 Best Local Similarity 27.2%; Pred. No. 3e-21;
 Matches 137; Conservative 77; Mismatches 174; Indels 116; Gaps 13;
 QY 10 TGMKTHSTGIGTSSNNIANANIGYQQQVQDLFSODLAIGSTGSCQPNAGKAQVIG 69
 Db 8 SGINAANADLNVTSSNNIANVNTTGFKESRAEPADMF-QSTSYGLS-----RNAVGSGRVS 62
 QY 70 SVRTITQGAPEPGNSVTDLAIGGKGFQVTEDEKVVHYTRAGNFRFTQDGLNDPSGFTL 129
 Db 63 NVAQQFSQGNIDPTGRSLDLAISGEGFAVSSNGAKMYTRAGNFQDANGYVINPQGARL 122
 QY 130 MGRISNNPN---IKKETLEPTQLDNDPTVAKSPAKTSTALNAVNL-GDSTDKTQSEA 185
 Db 123 --QVFAPNPSGNGFDVGRSLDLQLLTD-----SPPKSTSTVNLAFPLGNASAPVTTF 175
 QY 186 NPYFALLESMKNGTGPPISTSNYSYAO-PMRVYDQGNSHDITVYFDGAPSPSTGSKTPEY 244
 Db 176 N-----PADDKYSHSTGGINVDYSLGVSHVQTSYF-----206
 QY 245 LVAMNPSE-----DGSAASGTDSSAGLLMSGTMTFSSNGELKNMTAFTTTSATKDLNA 297
 Db 207 VKTANPNQVHNYVDGSAVGAPTTLQFSDTGLTTPANG-IVAMDPTPTP-----256
 QY 298 WQAPLVNGLPQFSANFVGAGIQPLTLDGFKSQNMWAGAPASAAAIGTDIGKLPSPMP 357
 Db 257 -----TGAGV-----LS 263
 QY 358 ICTSSGNSTARNGSSSTRYSODGYPQGLVDVITTSSEKLGKYSNSQVDFYNIPLAR 417
 Db 264 MQLNVSGSTQYGEAPALRDTROUGYASGKLEISIDTSGVWFARYSNGADKALGQVALNS 323
 QY 418 FTSEDGLRREGNNHYSATLDSGGPEFGLPGTSNYGKLSVNOLETNSVDMRSYMNMIIO 477
 Db 324 FVNPQGLSQGNMWAESYTSGAARVGAPDTSGLQIGSSLEASTVDLTQLVNMVIAQ 383
 QY 478 RGFQNMKSQVTTADTMLOKALELR 501
 Db 384 RNFQANSQWISTQDQVQTIIINR 407
 RESULT 38
 Q8PL23 PRELIMINARY; PRT; 407 AA.
 ID Q8PL23
 AC Q8PL23
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Flagellar biosynthesis hook protein.
 GN Name=flagE; OrderedLocusNames=XAC1983;
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RA MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

Query Match		18.6%; Score 481; DB 1; Length 718;
Best Local Similarity		24.8%; Pred. No. 6.7e-21;
Matches 178; Conservative		96; Mismatches 223; Indels 226; Gaps 21;
QY	1	MMGSLFIGATGKTHSTGLTGVSSNNIANANTIGYKQQQVVFQDLPSQDLAIGS-----TGS 56
DB	1	MLRSLWSGVNGMAHQIALDIESNNIANVNTTGFYKYSRASFDVMDLSQVKTATAPYKNGL 60
QY	57	QGNP--QAGMGQVGVRIFFTCGAFEPGNSVTDIAIGKGFQVTLDEKV--HYTRAGN 112
DB	61	AGNDPSVLGVGVDAATTKIFSCQNTQNTDVTDLAIQDGFIIISPRGITRNFTRDGE 120
QY	113	FRFTQDGLNDPSGFTLM-----GSRISNNPIKKTLEPIQLQDFNPVAKSPAK 163
DB	121	FUFDSQGLVTTGGLVVGWVNGSDTGKNGSDTDALKVDNTGPLENIRIDPGWV-MPAR 179
QY	164	TST--ALNAVNLGSDTKTQSEANPYFALLESWK--GNGTPISTSNYSYQA----- 212
DB	180	ASNRISMRLNAGRHADQTAA---IFALDSSAKTPSDGINPVYDSGTNLQAQVAEDMGS 235
QY	213	-----PMRVYDQDQ-----NSHDITVYFDGAP-----SSTGSKTFEY 244
DB	236	LCNEDGDALLINENQIIVSYSAKVVKDILPSAENSTLELNGVKISFTNDSAVSRSS 295
QY	245	LVAMPSEDSGAAG-----TDSAGLLMSGTMTFFSNGELKN----- 281
DB	296	VAAKNAINAVKSGTGEAYLDGKQLRLNTNLDGEXLKNIVVTQAGTGAFANFLDGD 355
QY	282	-MTAF-----TPTGSATKDLNAWQAPLVNGLPQPSANFV-----GAGIQPLTL 324
DB	356	DVTAPEKYSYTHSISNADIGOFRTTDLRALIQHDANIVKDPISLADNYQDSAAIGVSVN 415
QY	325	DFG-----IKSQQNNW-----AGAPASAA 343
DB	416	QXGMEFINNKNKNIKENLAFVSGYSSDSVTNNVLPKNAWKGLNTASLIEGASASS 475
QY	344	-----AIGT----- 347
DB	476	KPTHATHATSDIVSLGTHKAMRIEYRSGAEWFRVIVPEGELVGSAAFPNVFEG 535
QY	348	-----DIGKLPSMMP-----IQTSSGNSPARGSSSTRYS-----Q 379
DB	536	GRLHNNDGSLAGNPPILQFPKNGADAPORINLAFSGSGFDLTSVDKISETVAIEQ 595
QY	380	DGYQGDLDVVTITSEGLKQKYSNSQVDPYNTPLARFTSEDLRREGNNHYSATLD 439
DB	596	NGYQAGDLMVRFDSGDVLLGAFSGNRTLALAQAVALANFANDAGLQALGNGVFSQTGNSG 655
QY	440	GPEGLPCTSNYKGLSVNOLSTSNVDKSRMVNMIIRGFQMNKSVTTADTMLQKALE 499
DB	656	QALIGAANTGRGSGISGKLESSNVDLSRLTNLIVVQRFQANSKAVTTSDQILNTLN 715
QY	500	LKR 502
DB	716	LKQ 718
RESULT 40		
QY	Q7VFH2	PRELIMINARY; PRT; 718 AA.
ID	Q7VFH2	
AC	Q7VFH2	
DT	01-OCT-2003 (TrEMBLrel. 25, Created)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Flagellar hook protein FlgS.	
GN	Name=flgS; OrderedLocusNames=HH1704;	
OS	Helicobacter hepaticus.	
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;	
OC	Helicobacteriaceae; Helicobacter.	
OX	NCBI_TaxID=32025;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
STRAIN=ATCC 51449 / 3B1;		
MEDLINE=22709201; PubMed=12810954;		
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,		
RA Bell M., Droge M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,		
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,		
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.,		
ET "the complete genome sequence of the carcinogenic bacterium		
RT Helicobacter hepaticus";		
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).		
DR EMBL; AE017149; AAP78301.1; -		
DR GO; GO:0009288; C:flagellum (sensu Bacteria); IEA.		
DR GO; GO:0003774; F:motor activity; IEA.		
DR GO; GO:0005198; F:structural molecule activity; IEA.		
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.		
DR InterPro; IPR010930; DUF1078.		
DR InterPro; IPR011491; Flae.		
DR InterPro; IPR010810; Flagellin IN.		
DR InterPro; IPR001444; Flag bb rod.		
DR Pfam; PF06429; DUF1078; 1.		
DR Pfam; PF07559; Flae; 1.		
DR Pfam; PF07196; Flagellin IN; 1.		
DR Pfam; PF00460; Flg bb rod; 1.		
KW Complete proteome; Flagellum.		
SQ SEQUENCE : 718 AA; 77099 MW; 3938C46D468E4043 CRC64;		
Query Match		18.5%; Score 479; DB 2; Length 718;
Best Local Similarity		23.6%; Pred. No. 8.8e-21;
Matches 176; Conservative		90; Mismatches 209; Indels 270; Gaps 21;
QY	1	MMGSLFIGATGKTHSTGLTGVSSNNIANANTIGYKQQQVVFQDLPSQDLAIGST----- 54
DB	1	MLRSLWSGVNGMAHQIALDIESNNIANVNTTGFYKYSRASFDVMDLSQIKLATSYPKNGL 60
QY	55	GSQGNQAGMGQVGVRIFFTCGAFEPGNSVTDIAIGKGFQVTLDEKV--HYTRAGN 112
DB	61	GQNDPSVLGVGVDAATTKVFSQNTQNTDVTDLAIEGDFGFIISPRGTHNTRNGE 120
QY	113	FRFTQDGLNDPSGFTLMG-----SRISNNPIKKTLEPIQLD 153
DB	121	FUFDANGLVTTGGVVGWVPEPLEAAESGTMDSDFDFRVDNTGPNVNIQIDP----- 174
QY	154	DPTVAKSPAKTSTALNAVNLG-----DSTDKTQSEA----- 185
DB	175	-GMVPAKATKTTITURANLNAGRHDIDQDVAALDSTAQTAADGVVAIYDSRGVLTQVGE 233
QY	186	-----NPYFALLE-----SWK-----GNGTPISTSN----- 207
DB	234	DLGVLFNDGDAFALNENQIIVSYKTAARHETVTVNEVSTIGINGEKVSFSNNSAITG 293
QY	208	-----YSYAPQMRVYDQ----- 219
DB	294	VSSIVAAQNAINSLRDKTGTVNAVYDAGQLRIENRQMDGGKVKNRITSGTGVLQNFV 353
QY	220	QGNSHDITVY-----PDGAPSSGTS-KTEFYLAVM-----NBSE-----DGSAAAGT 260
DB	354	QG-EEDITAFRYRTKSEDASTTGQFRTTDLRALIQDANWIKNPEKTYESTATVGV 412
QY	261	-----DSAGLLMSGTMTFSSN----- 276
DB	413	TINRWGMEFIANHDDADDEQRNLSLYVTSFSDVNTNNVLPKTKMTKALNTASLIEGAAV 472
QY	277	-----GELKNMT-----AFTPTGSATKDLNAWQAPLVNGLPQPSANFV 315
DB	473	NTGKIVKATHATSVDIVDSLGSKKNVPEFVKTDGAVWSFRAIVPEP-----AQFI 523
QY	316	GAGIQPLTDFGIKQQNMWAGAPASAAAGTIDGKLPFMSVPIQTSNGS----- 365
DB	524	GGS1-----TKPNVEGGRASFNSDGLSGMNPVLPQDPKNGSKGPORLELKF 573
QY	366	-----TARNGSSSTRYSQDGPQGLVDVVTITSEGLKQKYSNSQVVDYFNIPAR 417
DB	574	ANETFGGLTSVDKIKSETYSINQYQAGDLMDFRDSNGSLLAGFSNGRSIALAQAVALAN 633

Search completed: October 26, 2004, 09:09:00
Job time : 206 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 31, 2004, 02:44:06 ; Search time 3026 Seconds
(without alignments)
6045.190 Million cell updates/sec

Title: US-10-009-823A-1
Perfect score: 2586
Sequence: 1 MMSLFTGATGKTHSTGLG.....NSKSVTTADTMLOKALELKR 502

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/US1009823/runat 26102004 100129 1617/app query.fasta_1.647
-DB=EST -OPT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=500 -THR_MIN=0 -THR_MAX=100 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1009823 @CGN_1_1 2607 @runat 26102004 100129 1617 -NCP=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOS=6 -DELEXT=7

Database : EST:

1: gb_est1:
2: gb_est2:
3: gb_hcc:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gsel:
9: gb_gsel2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	294	11.4	729	8 BH370667	BH370667 AG-ND-165
C 2	272	10.5	891	8 B2577315	B2577315 msh2.5350
C 3	221.5	8.6	1319	8 B2556688	B2556688 pasc1-60
C 4	221	8.5	742	8 BH795484	BH795484 23H03LLIO
C 5	214	8.3	1105	8 B2563024	B2563024 pasc2-164
C 6	209	8.1	752	7 CF891171	CF891171 UI-CF-DUI
C 7	208.5	8.1	895	8 B2577405	B2577405 msh2.5391
C 8	206.5	8.0	565	6 CB065273	CB065273 GGEZHT102
C 9	200.5	7.8	898	7 CO423013	CO423013 GGEZHT102

C 10	200.5	7.8	1190	8 BZ557649	BZ557649 pasc1-60
C 11	200	7.7	849	9 CL660824	CL660824 PRI0138a
C 12	199.5	7.7	824	7 CO202677	CO202677 Oa splbn
C 13	197.5	7.6	758	7 CO421506	CO421506 GGEZHT102
C 14	197.5	7.6	761	1 AJ747468	AJ747468 AJ747468
C 15	197.5	7.6	766	7 CO423511	CO423511 GGEZHT101
C 16	190.5	7.4	870	7 CN823216	CN823216 Oa splbn
C 17	183	7.1	650	6 CB854175	CB854175 UI-CF-DUI
C 18	181.5	7.0	661	7 CN582719	CN582719 USDA-FP-1
C 19	178	6.9	566	5 BP562788	BP562788 BP562788
C 20	175	6.8	436	6 CD119056	CD119056 ME1-0052U
C 21	175	6.8	436	6 CD119187	CD119187 ME1-0052U
C 22	172.5	6.7	532	1 AI057966	AI057966 SMOVL3CAN
C 23	169	6.5	715	5 CD444654	CD444654 EL01N0442
C 24	168	6.5	692	7 CF886932	CF886932 UI-CF-DUI
C 25	161.5	6.2	1158	8 BZ557676	BZ557676 pasc1-60
C 26	161.5	6.2	1208	8 BZ557650	BZ557650 pasc1-60
C 27	158.5	6.1	554	4 BI309317	BI309317 EST530727
C 28	154	6.0	747	7 CN822868	CN822868 Oa splbn
C 29	139	5.4	3057	3 AK044947	AK044947 Mus muscu
C 30	138	5.3	374	8 AZ577776	AZ577776 14f11 Sho
C 31	137.5	5.3	1590	9 AY412935	AY412935 Homo sapi
C 32	135.5	5.2	4428	9 AY416869	AY416869 Pan trogl
C 33	133.5	5.2	726	7 CN823464	CN823464 Oa splbn
C 34	133.5	5.2	4355	9 AY416870	AY416870 Mus muscu
C 35	132	5.1	374	7 CN762396	CN762396 IDOAAA4AB
C 36	131.5	5.1	348	4 BG857482	BG857482 1024052B0
C 37	131.5	5.1	1723	3 AK089973	AK089973 Mus muscu
C 38	128.5	5.0	324	8 AZ577764	AZ577764 14e10 Sho
C 39	128.5	5.0	4428	9 AY416868	AY416868 Homo sapi
C 40	128	4.9	867	8 BZ577319	BZ577319 msh2.5352
C 41	127.5	4.9	527	6 CA581013	CA581013 EST000688
C 42	126.5	4.9	882	8 BZ559876	BZ559876 pasc2-164
C 43	124.5	4.8	2293	3 AK032791	AK032791 Mus muscu
C 44	124.5	4.8	4722	3 AK029775	AK029775 Mus muscu
C 45	123	4.8	4489	3 AK028459	AK028459 Mus muscu

ALIGNMENTS

RESULT 1

BH370667/c

LOCUS BH370667

729 bp DNA linear GSS 10-DEC-2001

AG-ND-165M21.TF ND-TAM Anopheles gambiae genomic clone

AG-ND-165M21, genomic survey sequence.

ACCESSION BH370667

VERSION BH370667.1

KEYWORDS GI:17316792

SOURCE Anopheles gambiae (African malaria mosquito)

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

Anopheles.

1 (bases 1 to 729)

AUTHORS Hong, X.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J.,

Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B.,

Gardner, M.J., and Collins, F.H.

Construction of a BAC library and generation of BAC end

sequence-tagged connectors for genome sequencing of the African

malaria mosquito Anopheles gambiae

Mol. Genet. Genomics 268 (6), 720-728 (2003)

22542063

12655398

Other GSSs: AG-ND-165M21.TR

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by

F.H. Collins and sequenced by The Institute for Genomic Research

(TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.

Location/Qualifiers

1. .729
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-165M21"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

ORIGIN

Alignment Scores:
Pred. No.: 2,066-20 Length: 729
Score: 294.00 Matches: 729
Percent Similarity: 46.84% Conservative: 38
Best Local Similarity: 30.64% Mismatches: 77
Query Match: 11.37% Indels: 48
DB: 8 Gaps: 3

US-10-009-823A-1 (1-502) x BH370667 (1-729)

Qy 267 MetSerGlyThrMetThrPheSerSerAsnGlyLeuLeuLysAsnMetThrAlaPheThr 286
Db 727 ATGCCCGGTACCATGACGTTTCGACACCGGCAATCTGGCCAGCACCCACGACCAACAA 668
Qy 287 ProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGly 306
Db 667 GGCACCGCGCGGATTCAGATCTG----- 641
Qy 307 LeuProGlnPheSerAlaAsnPhaValGlyAlaGlyIleGlnProLeuThrLeuAspPhe 326
Db 640 ATCCCGATGAAGGCAAAAGAT-----GGCGCGCGGCACAGAACTTCACCTGGAATTC 587
Qy 327 GlyIleLysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGly 346
Db 586 GCCAAGACATGACGACGCGTC----- 563
Qy 347 ThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThr 366
Db 563 ----- 563
Qy 367 AlaArgAsnGlySerSerThrArgArgTySerGlnAspGlyTyProGlnGlyAsp 386
Db 562 -----AGCGCGGACTCCGTGACAAAGTCTCGCAGGATGGCTATGGCGGAGTAA 512
Qy 387 LeuValAspValThrIleThrSerGlyLysLeuGlnGlyLysTySerAsnSerGln 406
Db 511 TACACCACTTCCAGATCAACACGACGCGCAGCGTATGGGCACTTCTTCCAAACCGAG 452
Qy 407 ValValAspPheTyThrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArg 426
Db 451 ACTCAGGTGTCGGCGCAGATGTCATGGCCAACTTCTCCAAACCGAGGCGCTCTTCT 392
Qy 427 GluGlyAsnAsnHisTySerAlaThrLeuAspSerGlyGlyProLupheGlyLeuPro 446
Db 391 CAGGGCGATACCTTTGGGCAAGAACCGCGGATCCGCGCAGCCAGCGTAGCGCTGGCG 332
Qy 447 GlyThrSerAsnTyThrGlyLysLeuSerValAsnGlnLeuGlnThrSerAsnValAspMet 466
Db 331 GGCACCGGGGTTTGGGCAAACTGACACGCGGCGCTGGAAGCTTCCAAAGCTTCAATCTG 272
Qy 467 SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer 486
Db 271 AGCCAGGACTGTTGAACATGATCTGCGCACCAACGTAACCTACCGTGCAGACGCCCAACC 212

Qy 487 ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys 501
Db 211 ATCAAAACGACGACTCGATTCTGCAACGCTGTTAGCTGGCG 167

RESULT 2

BZ577315 891 bp DNA linear GSS 17-DEC-2002
LOCUS msh2_5350.x1 msh Pseudomonas aeruginosa genomic clone msh2_5350,
DEFINITION genomic survey sequence.

ACCESSION BZ577315

VERSION BZ577315.1 GI:27212376

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 891)

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

JOURNAL Pseudomonas aeruginosa library

COMMENT J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1. 891
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_5350"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."

ORIGIN

Alignment Scores:
Pred. No.: 6,166-18 Length: 891
Score: 272.00 Matches: 105
Percent Similarity: 30.06% Conservative: 45
Best Local Similarity: 21.04% Mismatches: 94
Query Match: 10.52% Indels: 255
DB: 8 Gaps: 11

US-10-009-823A-1 (1-502) x BZ577315 (1-891)

Qy 1 MetMetGlySerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGly 20
Db 93 ATGTTTCGGCAGTGTGGTTCAGCAAGACCGGTCTGTCCGCCCGGACATGACCTGACC 152
Qy 21 ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyThrLysGlnGlnValVal 40
Db 153 ACCATTTCACCAACACCTGGCCCAAGTATCCACCACCGCTTCAAGCGGACCGCGGAG 212
Qy 41 PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn 60
Db 213 TTCAGACCTGCTGTATCCAGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 272
Qy 61 GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla 79
Db 273 GAGCTGCTTCGGCGCTTCGCACTGGGTACCGGTGTGCGGTGTGCGCGCGCGCGCG 332
Qy 80 PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLys 94
Db 333 TTCACCCCGGCGCGCTGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392


```

Db      801 AACAGCGTGTGGACCTGCCCATCAACGGCAACGGCTTCTCGTCACCAAGCAAAACACGG 860
QY      104 LysValHisTyrThrArgAlaGlyAsnPheA:gpHeThrGlnAspGlyPheLeuAsnAsp 123
Db      861 GCGATCAGGTACACCGCGCGCTACTTCAT-TACCGACACGGGATTCATTCCGCGAC 919
QY      124 ProSerGlyPheThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLys--- 142
Db      920 CACAACCGTTACGCTGCGAGGCTTTGCGTCNGGCCGACCGCCAGTTGCAGAACNGGT 979
QY      143 -----GluThr 144
Db      980 TGCTAACGACCTTCAAGGTGAGCGGCCCAATAGCGCGGAGCACTGGGCATCCAAAGG 1039
QY      145 LeuGluProIleGlnLeuAspPheAsn-----AspProThrValAlaLysSerPro 161
Db      1040 GTACAACCT---AAATTGAACCTTGAACCCCGACGCGGACCCCTTTATCCTT-----CCC 1090
QY      162 AlaLysThrSerThrAlaLeuAsnAlaValValaenLeuGlyAspSerThrAspLysThr 181
Db      1091 CCGCGTACTAACTGGTCTC-----TTGGTTGGGATT-----1123
QY      182 GlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrPro 201
Db      1124 TTGAAGCGGTGCAATCCAACTTGGCTTTTATATAAAACACCGGTTCCTCAAGTTACCC 1183
QY      202 ProfileSer 204
Db      1184 CCGGTTTCG 1192

RESULT 4
BH795484/c
LOCUS
DEFINITION
23H03LLI0003B Lawsonia library DNA linear GSS 21-OCT-2002
Genomic survey sequence.
ACCESSION
BH795484
VERSION
BH795484.1 GI:24182434
KEYWORDS
GSS.
ORGANISM
Lawsonia intracellularis
Bacteria; Proteobacteria; Deltaproteobacteria; Desulfotribionales;
Desulfotribionaceae; Lawsonia.
REFERENCE
1 (bases 1 to 742)
Gebhart,C.J., Li,L., Zhang,Q., Guedes,R.M.C., Herron,L.L. and
Kapur,V.
Genomic Sequence Survey of Lawsonia intracellularis, the causative
agent of proliferative enteropathy, identifies numerous genes of
relevance to diagnosis, virulence, and immunopathogenesis
Unpublished (2002)
Contact: Zhang, Q
Department of Pathobiology
University of Minnesota
1971 Commonwealth Ave. St. Paul, MN 55108, USA
Tel: 612-625-9277
Fax: 612-625-5203
Email: qingzha@tc.umn.edu
Class: shotgun.
Location/Qualifiers
1. 742
/organism="Lawsonia intracellularis"
/mol_type="genomic DNA"
/db_xref="taxon:29546"
/clone_lib="Lawsonia library"
/note="Vector: pUC18; Site 1: SmaI"

FEATURES
source
Alignment Scores:
Pred. No.: 1.22e-12 Length: 742
Score: 221.00 Matches: 62
Percent Similarity: 48.97% Conservative: 33
Best Local Similarity: 31.96% Mismatches: 87
Query Match: 8.55% Indels: 12
DB: 8 Gaps: 3

ORIGIN

```

```

US-10-009-823A-1 (1-502) x BH795484 (1-742)
QY      316 GlyValaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrp 335
Db      682 GGAAGTCGCTCAACACCGTCCATAATTTTATCCAGGGGACTCACAGAAC-----632
QY      336 AlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMet 355
Db      631 ACAGTAATCTCTAATATGCTATTGAGGCAAGGGTTTTCACAGTA-----TTG 578
QY      356 MetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArg 375
Db      577 GTTGGTGATCAATTAATATATATACAGAGCAGGTGCTTTTAAACTTAATCAGGATGGACT 518
QY      376 ArgTyrSerGlnAspGlyTyrPro-----GlnGlyAspLeuVal 388
Db      517 CTGTACACGCCAATGATACCCCTCAACACAGAAATTTACAGTTCCTCTGATCTAAA 458
QY      389 AspValThrIleThrSerGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValVal 408
Db      457 TCTGTGCTATTTCTGAAAGAGGTCTGATTACTGCTTTAGACGCAATGGACAGAAATC 398
QY      409 AspPheTyrAsnIleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGly 428
Db      397 GCTGCTGGAGAAATACCTTTATATAGCTTTGTTAATCCAGCGGGACTCGTCAAAAGGA 338
QY      429 AsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThr 448
Db      337 CGTAATCTCTATATTTCCACAGACATCCGCTGAAGCGGTAGAGGTGCTCCCTGTGAA 278
QY      449 SerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArg 468
Db      277 GAAACTTAGGACATATAGCTCAAGGATTCCTTGAATGCTCTAATGTAGAGGTGTTGAT 218
QY      469 GluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThr 488
Db      217 GAAATGGTACAAATGATTGTAGGCAAAAGACATATGAATGAATTCNAAGCAATCCAA 158
QY      489 ThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLysArg 502
Db      157 ACTTCAGACAAATGCTTCAAACTGCAGTCCAACTTAAACGT 116

RESULT 5
BZ563024/c
LOCUS
DEFINITION
pacs2-164_405.sl pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_405, genomic survey sequence.
ACCESSION
BZ563024
VERSION
BZ563024.1 GI:27186093
KEYWORDS
GSS.
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1105)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. 1105
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"

FEATURES
source

```

```

/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164_405"
/clone_lib="pac2-164"
/note="clinical isolate 2-164 whole genomic shotgun
library."

ORIGIN
Alignment Scores:
Pred. No.: 1.26e-11 Length: 1105
Score: 214.00 Matches: 77
Percent Similarity: 47.03% Conservative: 34
Best Local Similarity: 32.63% Mismatches: 101
Query Match: 8.28% Indels: 24
DB: 7
Gaps: 7

US-10-009-823A-1 (1-502) x BZ563024 (1-1105)

QY 270 ThrMetThrPheSerSerAsnGlyGlu-----LeuLysAsnMetThrAlaPheThr 286
Db 1032 ACTTGCMAAATTTCCACAACTGGTTCAAGGAGACCGCTGGATTCAGGAACTTTTACC 973
QY 287 ProThrGlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGly 306
Db 972 AATACAGAACCAAGCGGGCGGCAATGTGACCATGACAGAGGAATCTCTCGCCCTGCAC 913
QY 307 LeuProGlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPhe 326
Db 912 TTGTTACCGTTTTCG-----GTGTCCTCGGA-----CCAGAGATCTTAACCC 868
QY 327 Gly---IleLysSerGlnGlnAsnMetThrAlaGlyAlaProAlaSerAlaAlaIle 345
Db 867 GGAAGCTGAAGACGACGACGAGCGCGTGG-----ACATTG 832
QY 346 GlyThrAspIleGlyLysLeuProSer-MetMet-----ProIleGlnThrSerSe 362
Db 831 GGGTAAAGGGCGGCTTTCTTCCAGGCTCTGTGTGTCAGCGCACCTCTCTACACCGCG 772
QY 362 rGlyAsnSerThr-----AlaArgAsnGlySerSerSerThrArgArgTyrSerGlnAs 380
Db 771 AGAAGCTCCACTTGAATCCGACGCGCAGATCGTCACCTTCCAAAGCTCGCCCTGGAGCCAG 712
QY 380 pGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlyGly-LysLeuGlnG 400
Db 711 CGATCGTGTGCCCATCGAGACCCAGACCTTCCACCGTCGGCGAGGACCGCTCTCTGTA 652
QY 400 lYlYsTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrS 420
Db 651 CCACACCGCGCAACCGCCGCGCGAGTGATCGGCACATCCACAGCCGCGACTCATCA 592
QY 420 exGluAspGlyLeuArgArgGlyGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyG 440
Db 591 ACCCGCGCGCTCGAGGCGCATCGGCAACACCTGTTCTCTGGAACCGGCTCCAGCGCG 532
QY 440 lYProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuG 460
Db 531 CGCCCGCAGGTCTGGTACGCGGGTCTCAACGCGCTCGGCAAGTTCGCCAGAACACCGCTG 472
QY 460 lUThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleIleGlnArgGlyP 480
Db 471 AAAACTCCAACTCAACGCTGGTGGAGAACTGGTGAACATGATCACCACCCAGCGGCT 412
QY 480 hGlnMetAsnSerLysSerValThrThrAlaAspThrMetLeu 494
Db 411 ACGAGATGAATCCCAAGGTGATCTTCCACCGCGCAGAGATGTTG 368

RESULT 6
CF8911171
LOCUS
DEFINITION
UI-CF-DUI-aas-f-05-18-UI.s18 UI-CF-DUI Homo sapiens cDNA clone
UI-CF-DUI-aas-f-05-18-UI 3', mRNA sequence.
CF8911171
ACCESSION
VERSION
CF8911171.1 GI:38150249

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 752)
Ronald, M.P., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/cf.html
Seq primer: M13 FORWARD
POLYA-No.

FEATURES
source

Location/Qualifiers
1..752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aas-f-05-18-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_hosts="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary lung epithelial cells The
library was constructed according to Ronald, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dr)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG_SEQ=None found"

ORIGIN

Alignment Scores:
Pred. No.: 2.34e-11 Length: 752
Score: 209.00 Matches: 46
Percent Similarity: 50.00% Conservative: 28
Best Local Similarity: 31.08% Mismatches: 70
Query Match: 8.08% Indels: 4
DB: 7
Gaps: 1

US-10-009-823A-1 (1-502) x CF891171 (1-752)

QY 357 ProfileGlnThr-----SerSerGlyAsnSerThrAlaArgAsnGlySerSer 372
Db 8 CCAACACAGACGCGGCTGCGCTCAACATGNGCGAACATCACCCAGTTCACCTCGGTACC 67
QY 373 SerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIle 392
Db 58 TACCGCAACCGCCGCCACCGACGCGTTCAGCCACCGGTCAATCACCAGCGGTGAATC 127
QY 393 ThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsn 412

Db	128	GACGGCAGCGGTGTGTTGTCGCACATTTCAGCAACACAGCAGACGAAGCCCATCGGCCAG	187
Qy	413	IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyr	432
Db	188	CTCTCCCTGGCCAGCCTTCAACAACAGCAGCAGGCGCTGCAGCCAGCGGGCGGCACCACTGG	247
Qy	433	SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly	452
Db	248	AAAGAAACCTTCGGCTCGGGCCAGCGGGTTTGATACCCCGCAAGTCGGACCCCTGGGT	307
Qy	453	LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn	472
Db	308	TCGATCTGGGCCAACTCCCTGGAGAACTCCAACCTCAACCTGACCAACGAGCTGCTGGAC	367
Qy	473	MetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr	492
Db	368	CTGATCAAGGCCCCAGAGCAACTATCAGCGCAACGCCAAGACCATCTCCACCCAAAGCACC	427
Qy	493	MetLeuGlnLysAlaLeuGluLeu	500
Db	428	ATCATGCAGACCATTCATCAGATG	451
RESULT	7		
LOCUS	BZ577405		
DEFINITION	msh2_5391.x1 msh Pseudomonas aeruginosa genomic clone msh2_5391, genomic survey sequence.		
ACCESSION	BZ577405		
VERSION	BZ577405.1		
KEYWORDS	GSS.		

233	ATGTTATCGGCACTGTGGGTGAGCAAGACCGGCTGTGTGCCCCAGAGCATGAACTGACC	2992	
Qy	21	ThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValVal	40
Db	293	ACCAATTCACAACAGCTGCCAAGCTATCCACCACCGCTTCAAGCGCAGCGCGGGAG	352
Qy	41	PheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn	60
Db	353	TTCAGGACCTCTGTACAGATCGGGCGCCAGCGGGCGCGCATCGACCCAGAGACGC	412
Qy	61	GlnAlaGlyMetGlyAlaGlnValGlySer---ValArgThrIlePheThrGlnGlyAla	79
Db	413	GAGCTGCCTTCGGGCGCTCAACTGGGTACCGGTGTGCGGTCTGCGCACCAGAGATC	472
Qy	80	PheGluProGlyAsnSerValThr-----AspLeuAlaIleGlyGlyLys	94
Db	473	TTCACCCGGGAGCGCTGCAGACACCCGAGACGCGCTGGACATCGCGTCAACGGGCGC	532
Qy	95	GlyPhePheGlnValThrLeuGluAspLys---ValHisTyrThrArgAlaGlyAsnPhe	113
Db	533	CGCTTCTTCCAGGTCCTGTGCGGGAGCGCACCGTGTCTACACCGCGAGCGCAGCTTT	592
Qy	114	ArgPheThrGlnAspGly	119
Db	593	CACCTGAACCTCGACGGG	610
RESULT 8			
CB065273			
LOCUS	EST644954	HOGA Medicago truncatula cDNA clone HOGA-1C13, mRNA	565 bp linear EST 21-JAN-2003
DEFINITION		sequence.	
ACCESSION	CB065273		
VERSION	CB065273.1	GI:27810851	
KEYWORDS	EST.		
SOURCE	Medicago truncatula (barrel medic)		
ORGANISM	Medicago truncatula		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.		
REFERENCE	1	(bases 1 to 565)	
AUTHORS	Hahn, M.G., Ojanen-Reuhs, T., Samac, D., Town, C.D., Van Aken, S., Utterback, T., Cho, J. and Fraser, C.M.		
TITLE	ESTs from roots of Medicago truncatula treated with oligogalacturonides of DP 6-20		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Michael G. Hahn Complex Carbohydrate Research Center University of Georgia 220 Riverbend Road, Athens, GA 30602-4712, USA Tel: 706-542-4457 Fax: 706-542-4412 Email: hahn@ccrc.uga.edu TIGR sequence name: MTMAAL9TK More information is available at: www.medicago.org Seq primer: SKmod (CTA GAA CTA gta gta CC). Location/Qualifiers 1..565 source		

```

ORIGIN
library."

Alignment Scores:
Pred. No.:      3,47e-11      Length:      895
Score:          208.50        Matches:      48
Percent Similarity: 55.56%    Conservative: 22
Best Local Similarity: 38.10% Mismatches:   49
Query Match:      8.06%      Indels:       7
DB:               8          Gaps:          3

US-10-009-823A-1 (1-502) x BZ577405 (1-895)

Oy      1 MetMetGlySerIenPheIleCtyAlaThrGlyMetIysThrHisSerThrGlyLeuGly 20
      |||:::  ::|||:::|||||:::

```

extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in SOLR cells."

ORIGIN

Alignment Scores:
Pred. No.: 2,75e-11 Length: 565
Score: 206.50 Matches: 47
Percent Similarity: 52.67% Conservative: 22
Best Local Similarity: 35.88% Mismatches: 55
Query Match: 7.99% Indels: 7
DB: 6 Gaps: 3

US-10-009-823A-1 (1-502) x CB065273 (1-565)

QY 6 PheileGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsn 25
Db 3 TGGGTCAGCAAGCGCGCTCGCGCTCATGACACCACTGACGACCATTCACCAAC 62
QY 26 lLeAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValPheGlnAspLeuPhe 45
Db 63 CTGGCCACGCTCTCGACGACCGGTTTCAAGAAGGACCGCGCGAGTTTCAGGACCTGCTG 122
QY 45 -----SerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsn--- 60
Db 123 TACCAGATCAAGCGCCACCGGTCGTCTAGTCACCAAGACGAGCGAGTCGCCACCGGT 182
QY 61 ---GlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAla 79
Db 183 CTGACGTCGCTACTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 242
QY 80 PheGluProGlyAsnSerValThrAspLeuAlaIleGlySerThrGlySerGlnGlyPheGlnVal 99
Db 243 CTGCAGACACCGACCGCGCTGGACATGGCCATCAACGGTCGGCGCTTCTTCAGATC 302
QY 100 ThrLeuGluAsp---LysValHisThrArgAlaGlyAsnPheArgPheThrGlnAsp 118
Db 303 ATCAGCCCGGACGACCATCTCTATACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 362
QY 119 GlyPheLeuAsnAspProSerGlyPheThrLeu 129
Db 363 GGTGAGTGTGTACCGCCCAACGGTCTGCGCGCTG 395

RESULT 9
CO423013 898 bp mRNA linear EST 02-JUL-2004
LOCUS GGEZHT1021A02.9 HT1 Gallus gallus cDNA clone GGEZHT1021A02, mRNA
DEFINITION sequence.

ACCESSION CO423013
VERSION CO423013.1 GI:496339261
KEYWORDS EST.
SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 898)
AUTHORS Silva, C.S., Jorge, E.C., Patricio, M., Ledur, M.C. and Coutinho, L.L.
TITLE Discovery of new genes expressed in the chicken pituitary and hypothalamus

JOURNAL Unpublished (2004)
COMMENT Contact: Clarissa S. Silva
Laboratory of Animal Biotechnology, Dep. of Animal Production
ESALQ - University of Sao Paulo
Av. Padua Dias, 11, Piracicaba, SP, 13418-900, Brazil
Tel: 55 19 3429 4434
Fax: 55 19 3429 4285
Email: ccsilva@esalq.usp.br and llcoutho@esalq.usp.br
PCR Primers
BACKWARD: T7.

FEATURES
Location/Qualifiers
1..898
/organism="Gallus gallus"

/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="GGEZHT1021A02"
/tissue_type="pituitary and hypothalamus"
/dev_stage="21 days old"
/lab_host="DH10B"
/clone_lib="HT1"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This cDNA library was constructed with the SuperScript Plasmid System with Gateway Technology kit (Invitrogen), following manufacturer's protocols. Plasmid DNA was purified using a modified alkaline lysis method. Sequencing reactions were conducted using the kit Big Dye Terminator Cycle Sequencing Ready Reaction (Applied Biosystems) according to the manufacturer's recommendations. Clones were sequenced by the 5' end with T7 primer. Sequencing reactions were analyzed on ABI Prism 3100 Genetic Analyzer (Applied Biosystems). The quality and clustering of the ESTs were analyzed using the softwares Phred/Cap3. Only EST sequences with Phred quality greater than 20 and at least 150 bp were considered for clustering."

ORIGIN

Alignment Scores:
Pred. No.: 2,47e-10 Length: 898
Score: 200.50 Matches: 50
Percent Similarity: 49.34% Conservative: 25
Best Local Similarity: 32.89% Mismatches: 70
Query Match: 7.75% Indels: 7
DB: 7 Gaps: 3

US-10-009-823A-1 (1-502) x CO423013 (1-898)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 431 TCTTTTAACTCGGCTTAGCGGCTCTATGCGGCCAACAAACAACTGGAGCTGACCGGC 490
QY 24 AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPheGlnAsp 43
Db 491 AACAACTCGCAACGTGCGGACCAACCGGCTTCAATCGTCCGCGCAATTCGCGAT 550
QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db 551 ATCTACGCG-----GCGTCAGCTGGCGCCGCCGACAGACAGCATCGGC 595
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
Db 596 AACGGGTAAACCTGGCGGAGTGTCCAGCAGTTCACCCAGGCTGACGTCAACACACAGC 655
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePhe---GlnValThrLeuGlu 102
Db 656 GCGCGCACGCTGACATGGCGATCCAGGCGCGCGCTTCTTCGTGCAAGAGGCGAGCGAC 715
QY 103 AspLysValHisThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122
Db 716 GGTTCGCTGGAGTACACCCGTAAACGTCCTTCGCGCGCGACAAAGACGGCTACATCAC 775
QY 123 AspProSerGlyPheThr---LeuMetGlySerArgIleSerAsnAsnProAsnIleLys 141
Db 776 AACAAACCGGCACCTCGCGCTGCAAGGCTACCGCGCGACGCGGACGCGCAACATCCAG 835
QY 142 LysGluThrLeuGluProIleGlnLeuAspPheAsn 153
Db 836 AAGGCGCGCTTGACGACCTTCAGCTCACTCTGCG 871

RESULT 10
BZ557649/c
LOCUS BZ557649.1
DEFINITION pacsl-60_645.s1 pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION BZ557649
VERSION BZ557649.1 GI:27170747
KEYWORDS GSS.

BZ557649 1190 bp DNA linear GSS 17-DEC-2002
pacsl-60_645.s1 pacsl-60 Pseudomonas aeruginosa genomic clone
BZ557649
BZ557649.1 GI:27170747
GSS.

SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1190)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source
1..1190
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."
ORIGIN
Alignment Scores:
Pred. No.: 3.84e-10 Length: 1190
Score: 200.50 Matches: 58
Percent Similarity: 53.42% Conservative: 28
Best Local Similarity: 36.02% Mismatches: 49
Query Match: 7.75% Indels: 26
DB: Gaps: 6
US-10-009-823A-1 (1-502) x BZ557649 (1-1190)
Qy 339 ProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeu-----ProSerMetMet 356
Db 504 CCGCGACGCGAGCTTCGACTGAACCTCCGACGGCAGATCGTCACCTCCACGGTTCG 445
Qy 357 ProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySer-SerSerThrArgAr 376
Db 444 CCT-----GGAGCCAGCGATCGTGGTCCCAACGACGACCCAGAC 406
Qy 376 gTyrSer-----GlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSe 394
Db 405 GTTCACCGTCGCGCAGGACGGC-----ACCGTGTGCGTGACAC 367
Qy 394 rGluglyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIlePr 414
Db 366 CACCGGCAACGCCAG-----CCGACAGGTGATCGGC-----AACATCCA 328
Qy 414 oLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAl 434
Db 327 GACCGCGACTTCATCAACCGCGCGCGCGGAAATCGATCGCTTGTCTCAACGCCCTCGGCACG 268
Qy 434 aThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLe 454
Db 267 AACCGGCTCCAGCGCGCGCGCGGAAATCGATCGCTTGTCTCAACGCCCTCGGCACG 208
Qy 454 uSerValAsnGlnLeuGlnThrSerAsnValAspMetSerArgGluMetValAsnMetIl 474
Db 207 TGCCGAGAACACCTGGAAACCTCAACGTCACGTCGAGGTCGTCGAGGAACTGGTGAACATGAT 148
Qy 474 eIleIleGlnArgGlyPheGlnMetAsnSerIysSerValThrAlaAspThrMetIle 494
Db 147 CACCACCCAGCGCGCTACGAGATGAACCTCAAGGTCACTCTCCACCGCGCGACCATGTT 88
Qy 494 u 494

Db 87 G 87
RESULT 11
CL660824/c
LOCUS
DEFINITION
PRI0138a.G01 - PRI0138a.B21 (849) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
CL660824
CL660824.1 GI:50146561
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 849)
AppADB: an AcedB database for the nematode satellite organism
Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
FEATURES
source
1..849
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"
ORIGIN
Alignment Scores:
Pred. No.: 2.56e-10 Length: 849
Score: 200.00 Matches: 40
Percent Similarity: 65.17% Conservative: 18
Best Local Similarity: 44.94% Mismatches: 31
Query Match: 7.73% Indels: 0
DB: Gaps: 0
US-10-009-823A-1 (1-502) x CL660824 (1-849)
Qy 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyr 432
Db 844 ATTGACTGGCGAATTTGCCAACACGAAGGCTGCATCCGAGGCGACACGCTCG 785
Qy 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 784 TCTGCGACGAATTTCTTGGCGTGGCGCTTTGGGACAGCCGGGAACTTTGGC 725
Qy 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 724 ACCCTGACACCGGTGGCGCTGGAGCGCTCCAAAGTCAATCTCAAGAGACTGGTCAAT 665
Qy 473 MetIleIleGlnArgGlyPheGlnMetAsnSerIysSerValThrAlaAspThr 492
Db 664 ATGATCGTTCCCGAGCTAACTATCATGCTTAAGCCGACCATCAAAACCCAGGCCAG 605
Qy 493 MetLeuGlnLysAlaLeuGluLeuLys 501
Db 604 ATCTCAACACGCTGGTTAACTTACGC 578

RESULT 12	
CO202677	
LOCUS	CO202677 824 bp mRNA linear EST 21-JUN-2004
DEFINITION	Oa_splbn_13E13_M13_Reverse Sheep spleen\brain pSport1 library Ovis aries cDNA clone Oa_splbn_13E13 5', mRNA sequence.
ACCESSION	CO202677
VERSION	CO202677.1 GI:49013852
KEYWORDS	EST.
SOURCE	Ovis aries (sheep)
ORGANISM	Ovis aries
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Caprinae; Ovis.
AUTHORS	1 (bases 1 to 824)
TITLE	Gossner, A. and Hopkins, J.
JOURNAL	Ovine spleen\brain cDNA library
COMMENT	Unpublished (2004)
CONTACT	Contact: J Hopkins
UNPUBLISHED	Veterinary Biomedical Sciences
UNIVERSITY	University of Edinburgh
ADDRESS	Summerhall Square, Edinburgh, EH9 1QH.
EMAIL	Email: j.hopkins@ed.ac.uk
PLATE	Plate: 13 row: E column: 13
SEQ PRIMER	Seq primer: M13 Reverse
High quality sequence stop: 548.	
FEATURES	Location/Qualifiers
source	1..824
	/organism="Ovis aries"
	/mol_type="mRNA"
	/db_xref="taxon:9940"
	/clone="Oa_splbn_13E13"
	/clone_lib="Sheep spleen\brain pSport1 library"
ORIGIN	
Alignment Scores:	
Pred. No.:	2,76e-10 Length: 824
Score:	199.50 Matches: 70
Percent Similarity:	38.08% Conservative: 21
Best Local Similarity:	29.29% Mismatches: 77
Query Match:	7.71% Indels: 71
DB:	7 Gaps: 8
US-10-009-823A-1 (1-502) x CO202677 (1-824)	
Qy	298 TrpGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnValGlyAla 317
Db	1 TGGTACCTCGCA-----GTACCGGTC 21
Qy	318 GlyTleGlnProLeuThrLeuAspPheGlyIleIysSerGlnGlnAsnMetTrpAlaGly 337
Db	22 CGGATTCGGGTGCGAGCCAGGACAGCAATTCGGCTCGGGCTGCACTGGGTACCGGT 81
Qy	338 AlaProAlaSerAlaAlaIleGlyThr-----AspTleGlyLysLeuPro 353
Db	82 -----GTGGCATCGTTGGCACTCAGAGAACTTCAGCGCCGTAACCTTCAG 129
Qy	354 SerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySer----- 371
Db	130 -----CAACCGCCAGCCCGCTGACTTGGCGGTTCACGCTAAGGGTTTC 174
Qy	372 -----SerSerThrArgArgTySerGlnAspGly----- 381
Db	175 TTCAGATCTCTGACGCGGACGCGACACCACTACACCGCGACGCTAGTTCACCTG 234
Qy	381 ----- 381
Db	235 GACGCCAATGCCAGGTCTGTACCGCAACAGGTTTCGCGTGGAGCGCGGATGTGGTG 294
Qy	382 -----TyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
Db	295 CCGAACACGCCAGACTTTCACCGTGGCAACGACGCGCACCGTGTGATCATCTGTGGCC 354
Qy	396 GlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu 415

Db	355 GGCAAC-----CCGGCCTCCAGGTGATCGGC-----AACCTGCAAAACC 393
Qy	416 AlaArgPheThrSerGluAspGlyLeuArgArgGluGlyValAsnAsnHisTyrSerAlaThr 435
Db	394 GCCGACTTCATCAACCCCGCGGCTGCGAGGATTCGTAAACACCTGTTCTCTGGAACC 453
Qy	436 LeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSer 455
Db	454 GCGTCCAGCGCGCGCAAAATCGGCACCCCTGGCTTAACGGTTTGGTACCACCTG 513
Qy	456 ValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIle 475
Db	514 CAAGACACCCCTGGAACGTCACACGTCAGCACCGCTTGAGAGAGATGGTCAACATGATCACC 573
Qy	476 IleGlnArgGlyPheGlnMetAsnSerIysSerValThrThrAlaAspThrMetLeu 494
Db	574 ACTCAGCGCGCTACGAGATGAACCTCCAGGTGATTTCCACCGCGGACCATGCTT 630
RESULT 13	
LOCUS	CO421506 758 bp mRNA linear EST 02-JUL-2004
DEFINITION	GGEZHC1025H08.g Hc1 Gallus gallus cDNA clone GGEZHC1025H08, mRNA sequence.
ACCESSION	CO421506.1 GI:49637754
VERSION	CO421506
KEYWORDS	EST.
SOURCE	Gallus gallus (chicken)
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS	1 (bases 1 to 758)
TITLE	Silva, C.S., Jorge, E.C., Patricio, M., Ledur, M.C. and Coutinho, L.L. Discovery of new genes expressed in the chicken pituitary and hypothalamus
JOURNAL	Unpublished (2004)
COMMENT	Contact: Clarissa S. Silva Laboratory of Animal Biotechnology, Dep. of Animal Production ESALQ - University of Sao Paulo Av. Pádua Dias, 11, Piracicaba, SP, 13418-900, Brazil Tel: 55 19 3429 4434 Fax: 55 19 3429 4285 Email: cssilva@esalq.usp.br and llcouthin@esalq.usp.br PCR Primers: BACKWARD: 17.
FEATURES	Location/Qualifiers
source	1..758
	/organism="Gallus gallus"
	/mol_type="mRNA"
	/db_xref="taxon:9031"
	/clone="GGEZHC1025H08"
	/tissue_type="pituitary and hypothalamus"
	/dev_stage="21 days old"
	/lab_host="DH10B"
	/clone_lib="HC1"
	/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This cDNA library was constructed with the SuperScript plasmid System with Gateway Technology kit (Invitrogen), following manufacture's protocols. Plasmid DNA was purified using a modified alkaline lysis method. Sequencing reactions were conducted using the kit Big Dye Terminator Cycle Sequencing Ready Reaction (Applied Biosystems) according to the manufacturer's recommendations. Clones were sequenced by the 5' end with T7 primer. Sequencing reactions were analyzed on ABI Prism 3100 Genetic Analyzer (Applied Biosystems). The quality and clustering of the ESTs were analyzed using the software Phred/Cap3. Only EST sequences with Phred quality greater than 20 and at least 150 bp were considered for clustering."
ORIGIN	
Alignment Scores:	

Pred. No.: 3.95e-10 Length: 758
 Score: 197.50 Matches: 65
 Percent Similarity: 41.10% Conservative: 25
 Best Local Similarity: 29.68% Mismatches: 84
 Query Match: 7.64% Indels: 45
 DB: 7 Gaps: 7

US-10-009-823A-1 (1-502) x C0421506 (1-758)

QY 299 GlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGly 318
 DB 2 CAGGACAGCAATTGCGCTGGCGCTGCA-----CTGGGTACCGGT 43
 QY 319 IleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAla 338
 DB 44 GTGGCCATCGTTGGCACTCAGAGAAGACTTCAGCCCGGTAACTTGCAGCAAAACCGGCCAG 103
 QY 339 ProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIle 358
 DB 104 CCGTGGACTTGGGGTCAACGGTAAG-----GGTTCTTCAGATCCCTGCAGCCGAC 157
 QY 359 GlnThr-----SerSerGlyAsnSerThr 366
 DB 158 GGCACCACTACACCGCGCGGTACGTTCCACCTGGACGCCAATGCCAGTGGTGGT 217
 QY 367 AlaArgAsnGly-----SerSerSerThrArg 375
 DB 218 ACCGCCACGGTTTCGCGCTGGACCGCGGAGTGTGGTGGCAACACGCCAGACTTTC 277
 QY 376 ArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
 DB 278 ACCGTGGCAACGACGGC-----ACCGTGTGATCACTGTGGCC 316
 QY 396 GlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu 415
 DB 317 GGCAC-----CCGGCTCCAGGTGATCGGC-----AACCTGCAAAACC 355
 QY 416 AlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThr 435
 DB 356 GCCGACTTCATCAACCGCGCGCTCGACGGGATGTGTAACACCTGTTCTCTGGAAACC 415
 QY 436 LeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSer 455
 DB 416 GCGTCCAGCGCGCGCAATCGGACCCCTTAAACGGTTTGTGTACCCCTG 475
 QY 456 ValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIle 475
 DB 476 CAAGACCCCTGGAAAGCTCCAGCTCAGCACCGTTGAAGAGATGTTCAACATGATCACC 535
 QY 476 IleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMetLeu 494
 DB 536 ACTCAGCGCGCTACGAGATGAATCCAGGTGATTTCCACCGCGCGCAGATGCTT 592

RESULT 14
 AJ747468 761 bp mRNA linear EST 07-JUL-2004
 LOCUS
 DEFINITION AJ747468 forward - stimulated minus unstimulated macrophage Sus
 scrofa cDNA clone F_C0001825b_F11, mRNA sequence.

ACCESSION AJ747468
 VERSION AJ747468.1 GI:49917717
 KEYWORDS EST.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 761)
 AUTHORS Hopwood, P.A., Zhang, F., Lowden, S., Talbot, R., Burt, D., Archibald, A.
 TITLE Development of a porcine cDNA microarray
 JOURNAL Unpublished (2004)
 COMMENT Contact: Hopwood PA
 Dept. of Preclinical Veterinary Sciences
 Royal School for Veterinary Studies

Summerhall, Edinburgh, EH9 1QH, UNITED KINGDOM
 Sequencing was performed by ARK Genomics. This clone is available
 from ARK- Genomics, Roslin Institute, Roslin, Midlothian EH25 9PS,
 UK. See www.ark-genomics.org or contact info@arkgenomics.org.

FEATURES

Location/Qualifiers
 source
 1..761
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /clone="F_C0001825b_F11"
 /tissue_type="lung"
 /cell_type="macrophage"
 /clone_lib="forward - stimulated minus unstimulated
 macrophage"

ORIGIN

Alignment Scores:
 Pred. No.: 3.97e-10 Length: 761
 Score: 197.50 Matches: 47
 Percent Similarity: 57.50% Conservative: 22
 Best Local Similarity: 39.17% Mismatches: 44
 Query Match: 7.64% Indels: 7
 DB: 1 Gaps: 2

US-10-009-823A-1 (1-502) x AJ747468 (1-761)

QY 385 GlyAspLeuValAspValThrIleThrSerGluGlyLysLeu-----GlnGly 400
 DB 60 GCGGAATCGTTAAGTTCACCATCGTCTGATGCGTGGTACCGGTAAACCAACAGGC 119
 QY 401 LysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAlaArgPheThrSer 420
 DB 120 CAGGACGCTCCGTTTCAGTT-----GGGACGCTCAATCTCACCACCTTTATGAAT 170
 QY 421 GluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGly 440
 DB 171 GACACCGGGCTGGAGAGCATTTGGGAAACCTCTACACCGAAACGCAATCTCTGGTGCA 230
 QY 441 ProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGlu 460
 DB 231 CGAAGCAAGACACCGCGGGCTGACCGCGCGGACTGCTGTATCAAGGTATGTTGAA 290
 QY 461 ThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGlyPhe 480
 DB 291 ACCTCTAAGCTCAACGTGGCGAAGAACTGCTCAATATGATTACAGTGCACCGGCTTAC 350
 QY 481 GlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAlaLeuGluLeu 500
 DB 351 GAAATCAACAGTAAAGCGTGTCCACCACCATCAGATGCTGCAAAACTGACCGCACTC 410

RESULT 15
 C0423511 766 bp mRNA linear EST 02-JUL-2004
 LOCUS
 DEFINITION GGEZHT1017H04.g HT1 Gallus gallus cDNA clone GGEZHT1017H04, mRNA
 sequence.

ACCESSION C0423511
 VERSION C0423511.1 GI:49639759
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 766)
 AUTHORS Silva, C.S., Jorge, E.C., Patricio, M., Ledur, M.C. and Coutinho, L.L.
 TITLE Discovery of new genes expressed in the chicken pituitary and
 hypothalamus
 JOURNAL Unpublished (2004)
 COMMENT Contact: Clarissa S. Silva
 Laboratory of Animal Biotechnology, Dep. of Animal Production
 ESAU - University of Sao Paulo
 Av. Padua Dias, 11, Piracicaba, SP, 13418-900, Brazil
 Tel: 55 19 3429 4434

Fax: 55 19 3429 4285
Email: cssilva@aalq.usp.br and llcoutine@aalq.usp.br
PCR Primers
BACKWARD: T7.

FEATURES	Location/Qualifiers
source	1. .766

```

1: 708
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="GGEZHT1017H04"
/tissue_type="pituitary a
/dev_stage="21 days old"
/lab_host="DH10B"
1: 709
1: 710
1: 711
1: 712
1: 713
1: 714
1: 715
1: 716
1: 717
1: 718
1: 719
1: 720
1: 721
1: 722
1: 723
1: 724
1: 725
1: 726
1: 727
1: 728
1: 729
1: 730
1: 731
1: 732
1: 733
1: 734
1: 735
1: 736
1: 737
1: 738
1: 739
1: 740
1: 741
1: 742
1: 743
1: 744
1: 745
1: 746
1: 747
1: 748
1: 749
1: 750
1: 751
1: 752
1: 753
1: 754
1: 755
1: 756
1: 757
1: 758
1: 759
1: 760
1: 761
1: 762
1: 763
1: 764
1: 765
1: 766
1: 767
1: 768
1: 769
1: 770
1: 771
1: 772
1: 773
1: 774
1: 775
1: 776
1: 777
1: 778
1: 779
1: 780
1: 781
1: 782
1: 783
1: 784
1: 785
1: 786
1: 787
1: 788
1: 789
1: 790
1: 791
1: 792
1: 793
1: 794
1: 795
1: 796
1: 797
1: 798
1: 799
1: 800
1: 801
1: 802
1: 803
1: 804
1: 805
1: 806
1: 807
1: 808
1: 809
1: 810
1: 811
1: 812
1: 813
1: 814
1: 815
1: 816
1: 817
1: 818
1: 819
1: 820
1: 821
1: 822
1: 823
1: 824
1: 825
1: 826
1: 827
1: 828
1: 829
1: 830
1: 831
1: 832
1: 833
1: 834
1: 835
1: 836
1: 837
1: 838
1: 839
1: 840
1: 841
1: 842
1: 843
1: 844
1: 845
1: 846
1: 847
1: 848
1: 849
1: 850
1: 851
1: 852
1: 853
1: 854
1: 855
1: 856
1: 857
1: 858
1: 859
1: 860
1: 861
1: 862
1: 863
1: 864
1: 865
1: 866
1: 867
1: 868
1: 869
1: 870
1: 871
1: 872
1: 873
1: 874
1: 875
1: 876
1: 877
1: 878
1: 879
1: 880
1: 881
1: 882
1: 883
1: 884
1: 885
1: 886
1: 887
1: 888
1: 889
1: 890
1: 891
1: 892
1: 893
1: 894
1: 895
1: 896
1: 897
1: 898
1: 899
1: 900
1: 901
1: 902
1: 903
1: 904
1: 905
1: 906
1: 907
1: 908
1: 909
1: 910
1: 911
1: 912
1: 913
1: 914
1: 915
1: 916
1: 917
1: 918
1: 919
1: 920
1: 921
1: 922
1: 923
1: 924
1: 925
1: 926
1: 927
1: 928
1: 929
1: 930
1: 931
1: 932
1: 933
1: 934
1: 935
1: 936
1: 937
1: 938
1: 939
1: 940
1: 941
1: 942
1: 943
1: 944
1: 945
1: 946
1: 947
1: 948
1: 949
1: 950
1: 951
1: 952
1: 953
1: 954
1: 955
1: 956
1: 957
1: 958
1: 959
1: 960
1: 961
1: 962
1: 963
1: 964
1: 965
1: 966
1: 967
1: 968
1: 969
1: 970
1: 971
1: 972
1: 973
1: 974
1: 975
1: 976
1: 977
1: 978
1: 979
1: 980
1: 981
1: 982
1: 983
1: 984
1: 985
1: 986
1: 987
1: 988
1: 989
1: 990
1: 991
1: 992
1: 993
1: 994
1: 995
1: 996
1: 997
1: 998
1: 999
1: 1000
1: 1001
1: 1002
1: 1003
1: 1004
1: 1005
1: 1006
1: 1007
1: 1008
1: 1009
1: 1010
1: 1011
1: 1012
1: 1013
1: 1014
1: 1015
1: 1016
1: 1017
1: 1018
1: 1019
1: 1020
1: 1021
1: 1022
1: 1023
1: 1024
1: 1025
1: 1026
1: 1027
1: 1028
1: 1029
1: 1030
1: 1031
1: 1032
1: 1033
1: 1034
1: 1035
1: 1036
1: 1037
1: 1038
1: 1039
1: 1040
1: 1041
1: 1042
1: 1043
1: 1044
1: 1045
1: 1046
1: 1047
1: 1048
1: 1049
1: 1050
1: 1051
1: 1052
1: 1053
1: 1054
1: 1055
1: 1056
1: 1057
1: 1058
1: 1059
1: 1060
1: 1061
1: 1062
1: 1063
1: 1064
1: 1065
1: 1066
1: 1067
1: 1068
1: 1069
1: 1070
1: 1071
1: 1072
1: 1073
1: 1074
1: 1075
1: 1076
1: 1077
1: 1078
1: 1079
1: 1080
1: 1081
1: 1082
1: 1083
1: 1084
1: 1085
1: 1086
1: 1087
1: 1088
1: 1089
1: 1090
1: 1091
1: 1092
1: 1093
1: 1094
1: 1095
1: 1096
1: 1097
1: 1098
1: 1099
1: 1100
1: 1101
1: 1102
1: 1103
1: 1104
1: 1105
1: 1106
1: 1107
1: 1108
1: 1109
1: 1110
1: 1111
1: 1112
1: 1113
1: 1114
1: 1115
1: 1116
1: 1117
1: 1118
1: 1119
1: 1120
1: 1121
1: 1122
1: 1123
1: 1124
1: 1125
1: 1126
1: 1127
1: 1128
1: 1129
1: 1130
1: 1131
1: 1132
1: 1133
1: 1134
1: 1135
1: 1136
1: 1137
1: 1138
1: 1139
1: 1140
1: 1141
1: 1142
1: 1143
1: 1144
1: 1145
1: 1146
1: 1147
1: 1148
1: 1149
1: 1150
1: 1151
1: 1152
1: 1153
1: 1154
1: 1155
1: 1156
1: 1157
1: 1158
1: 1159
1: 1160
1: 1161
1: 1162
1: 1163
1: 1164
1: 1165
1: 1166
1: 1167
1: 1168
1: 1169
1: 1170
1: 1171
1: 1172
1: 1173
1: 1174
1: 1175
1: 1176
1: 1177
1: 1178
1: 1179
1: 1180
1: 1181
1: 1182
1: 1183
1: 1184
1: 1185
1: 1186
1: 1187
1: 1188
1: 1189
1: 1190
1: 1191
1: 1192
1: 1193
1: 1194
1: 1195
1: 1196
1: 1197
1: 1198
1: 1199
1: 1200
1: 1201
1: 1202
1: 1203
1: 1204
1: 1205
1: 1206
1: 1207
1: 1208
1: 1209
1: 1210
1: 1211
1: 1212
1: 1213
1: 1214
1: 1215
1: 1216
1: 1217
1: 1218
1: 1219
1: 1220
1: 1221
1: 1222
1: 1223
1: 1224
1: 1225
1: 1226
1: 1227
1: 1228
1: 1229
1: 1230
1: 1231
1: 1232
1: 1233
1: 1234
1: 1235
1: 1236
1: 1237
1: 1238
1: 1239
1: 1240
1: 1241
1: 1242
1: 1243
1: 1244
1: 1245
1: 1246
```

/clone_1105-n11.
NotI; Site 2: Salt; This
cDNA library was constructed with the Superscript Plasmid
System with Gateway Technology kit (Invitrogen), following
manufacturer's protocols. Plasmid DNA was purified using a
modified alkaline lysis method. Sequencing reactions were
conducted using the kit Big Dye Terminator Cycle
Sequencing Ready Reaction (Applied Biosystems) according
to the manufacturer's recommendations. Clones were
sequenced by the 5' end with T7 primer. Sequencing
reactions were analyzed on ABI Prism 3100 Genetic Analyzer
(Applied Biosystems). The quality and clustering of the
ESTs were analyzed using the softwares Phred/Cap3. Only
EST sequences with Phred quality greater than 20 and at
least 150 bp were considered for clustering."

ORIGIN

Alignment Scores:	4.01e-10	Length:	766
Pred. No.:	Score:	Matches:	65
	197.50	Conservative:	25
Percent Similarity:	41.10%	Mismatches:	84
Best Local Similarity:	29.68%	Indels:	45
Query Match:	7.64%	Gaps:	7
DB:	7		

U.S.-10-009-823A-1 (1-502) x CO423511 (1-766)

	Qy	299	GlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGly	318
	Db	2	CAGGACAGCGAATTGCCTCGGGCTGCAA-----CTGGGTACC GGTT 43	
	Qy	319	IleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTrpAlaGlyAla	338
	Db	44	GTGGCCATCGTGGCAGCTCAGAAAGACTTCAGCGCGGTAACTTGCAGCAAAACGGGCCAG 103	
	Qy	339	ProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMetProIle	358
	Db	104	CCGCTGGACTTGCGGCTCAACGGTAAG-----GGTTTCTTCCAGATCCTCGACGCCGAC 157	
	Qy	359	GlnThr-----SerSerGlyVAsnSerThr 366	
	Db	158	GGCACACACACCTACACCCGCGACGGTAGCTTCCACTGGAGCGCAATGGCCAGGTCGTG 217	
	Qy	367	AlaArgAsnGly-----SerSerSerThrArg 375	
	Db	218	ACGCCAACGGTTTCGCGCTGGAGCGCGCATTTGGTGGCGGAACAACAGCCGACATTTTC 277	
	Qy	376	ArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu	395
	Db	278	ACCGTCGGCAACACCGGC-----ACCGTGTGATCATCATCTGTGGCC 316	
	Qy	396	GlyLysLeuGlnGlyIlyTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu	415
	Db	317	GGCAAC-----CGGGCTCCCAGGTGATCGGC-----AACCTGCMAACC 355	
	Qy	416	AlaArgPheThrSerGlnAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThr	435
	Db	356	GCGACTTCATCAACCCCGCGCTCGGCGGATTTGGTAACAACCTGTTCCTGGAAACC 415	
	Qy	436	LeuAsoSerGlyGlvProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSer	455

[illegible]

	EST	EST ID	Accession	Gene	Species	Library	mRNA	bp	CN823216
LOCUS								870 bp	
DEFINITION									
				Oa splbn	M13reverse	Sheep spleen\brain pSport1 library			
				Oa splbn	04023	5'	mRNA sequence.		

ACCESSION	CN823216
VERSION	CN823216.1
KEYWORDS	EST

SOURCE
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Caprinae; Ovis.

REFERENCE 1 (bases 1 to 870)

AUTHORS Gossner, A. and Hopkins, J.

TITLE Ovine spleen/brain cdna library
JOURNAL Unpublished (2004)
COMMENT Contact: J Hopkins

Veterinary Biomedical Sciences

University of Edinburgh

Summerhall Square, Edinburgh, EH9 1QH.

Email: j.hopkins@ed.ac.uk

Plate: 04 row: 0 column: 23

seq primer: M13reverse

High quality sequence	start	end	length	score
1	1	100	100	100
2	1	100	100	100
3	1	100	100	100
4	1	100	100	100
5	1	100	100	100
6	1	100	100	100
7	1	100	100	100
8	1	100	100	100
9	1	100	100	100
10	1	100	100	100
11	1	100	100	100
12	1	100	100	100
13	1	100	100	100
14	1	100	100	100
15	1	100	100	100
16	1	100	100	100
17	1	100	100	100
18	1	100	100	100
19	1	100	100	100
20	1	100	100	100
21	1	100	100	100
22	1	100	100	100
23	1	100	100	100
24	1	100	100	100
25	1	100	100	100
26	1	100	100	100
27	1	100	100	100
28	1	100	100	100
29	1	100	100	100
30	1	100	100	100
31	1	100	100	100
32	1	100	100	100
33	1	100	100	100
34	1	100	100	100
35	1	100	100	100
36	1	100	100	100
37	1	100	100	100
38	1	100	100	100
39	1	100	100	100
40	1	100	100	100
41	1	100	100	100
42	1	100	100	100
43	1	100	100	100
44	1	100	100	100
45	1	100	100	100
46	1	100	100	100
47	1	100	100	100
48	1	100	100	100
49	1	100	100	100
50	1	100	100	100
51	1	100	100	100
52	1	100	100	100
53	1	100	100	100
54	1	100	100	100
55	1	100	100	100
56	1	100	100	100
57	1	100	100	100
58	1	100	100	100
59	1	100	100	100
60	1	100	100	100
61	1	100	100	100
62	1	100	100	100
63	1	100	100	100
64	1	100	100	100
65	1	100	100	100
66	1	100	100	100
67	1	100	100	100
68	1	100	100	100
69	1	100	100	100
70	1	100	100	100
71	1	100	100	100
72	1	100	100	100
73	1	100	100	100
74	1	100	100	100
75	1	100	100	100
76	1	100	100	100
77	1	100	100	100
78	1	100	100	100
79	1	100	100	100
80	1	100	100	100
81	1	100	100	100
82	1	100	100	100
83	1	100		

High quality sequence stop:
 Location/Qualifiers

FEATURES	LOCATION
1	870

Source
T. 870
Organism—"Ovis aries"

```
/original=0.15  
/mol type="mRNA"
```

```
/db xref="taxon:9940"
```

```
/clone="Oa splbn 04023"
```

```
/clone lib="Sheep spleen\brain pSport1 library"
```

ORIGIN

Alignment Scores:	2.72e-09	Length:	870
Pred. No.:	190.50	Matches:	50
Score:	48.00%	Conservative:	22
Percent Similarity:	33.33%	Mismatches:	71
Best Local Similarity:	7.37%	Indels:	7
Query Match:	7	Gaps:	3
DB:			

US-10-008-833A-1 (1-502) x CN823216 (1-870)

QY	4	SerLeuPheIleGlyAlaThrGlyMetIysThrHisSerThrGlyLeuGlyThrValSer	23
DB	433	TCCTTTAAATCGGGCTTAGCGGCTCTATCGCGCCCAACAAACAACTGGACGTACCGGC	492
QY	24	AsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnGlnValValPheGlnAsp	43
DB	493	AACAACATCGCAACGTCGCGACCAACCGGGCTTCAATCTCCCGTGCAGAAATTCCTCGGAT	552
QY	44	LeuPheSerClnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly	63
DB	553	ATCTACGG-----CGCTCCAAGCTGGCCACCGCCAGAACACATCGGC	597
QY	64	MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly	83
DB	598	AACGGCGTAAACCTCGCGCCAGTGTCCACGAGATTCAACAGGTGACGTCAACAACAGC	657
QY	84	AsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePhe---GlnValThrLeuGlu	102

```

Db 658 GSCGCGACCTGGACATCGCATCCAGGCGCGGCTCTTCGTGCGAGAGGGGCGAGC 717
      |||:::||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 103 AsplysvalHistyThrArgAlaGlyAsnPhaArgPheThrGlnAspGlyPheLeuAsn 122
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 GGTTCGCTGGAGTACACCGCTAACGGTTCCTTCGCGCGGCGGCTAGAGCGGTACATCACC 777
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 123 AspProSerGlyPheThr---LeuMetGlySerArgIleSerAsnProAsnIleLys 141
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 AACACACGGGACCTCGCGCTGACAGGCTACGCGCGGACCGGACCGGACATCCAG 837
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 142 LysGluThrLeuGluProIleGlnLeuAsp 151
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 AAGGGCGCTTGACCGCGACTTCAGCTCAAC 867
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 17
CB854175 650 bp mRNA linear EST 22-APR-2003
LOCUS UI-CF-DUI-aas-f-05-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
DEFINITION UI-CF-DUI-aas-f-05-0-UI 3', mRNA sequence.
ACCESSION CB854175
VERSION CB854175.1 GI:30044552
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 650)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (3), 791-806 (1996)
MEDLINE 97044477
PUBMED 889548
COMMENT Contact: McGray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mcgrayuiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=No.

```

```

FEATURES
source
1..650
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aas-f-05-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dr)18 tail. The sequence tag for this

```

```

library is GGCTGTAGGC.
TAG_SEQ=None found"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1.08e-08 Length: 650
Score: 183.00 Matches: 46
Percent Similarity: 49.33% Conservative: 28
Best Local Similarity: 30.67% Mismatches: 70
Indels: 6
Query Match: 7.08% Gaps: 1
DB: 6

```

```

US-10-009-823A-1 (1-502) x CB854175 (1-650)

```

```

Qy 357 ProIleGlnThr-----SerSerGlyAsnSerThrAlaArgAsnGlySerSer 372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 CCACACAGACGGCGCGTGCCTCAACATGNGCGGAACATCACCAGTTCACCTCGGTACC 68
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 373 SerThrArgArgTyxSerGlnAspGlyTyxProGlnGlyAspLeuValAspValThrIle 392
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 TACCGCAACCGCGCCACCACCGAGCGTTTACGCCACCGCTCAAAATCACCGGCGCTGAAAATC 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 393 ThrSerGluGlyLysLeuGlnGlyLysTyxSerAsnSerGlnValValAspPheTyxAsn 412
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 GAGCGACGCGGTGTGTGTTGTCGCCACTTTCAGCAACCCAGCAGCAAGCGGCATCGGCCAG 188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisty 432
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 CTCCTCCCTGGCCAGCTTCAACACGACGAGCGGCTCGAGCCGCGGCGGACCATCTGG 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 433 -SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyx 452
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 GAAAGAAACCTTCGCGTGGCGCGCGGCTTTGTATACCCCGCAAGTCGCGGACCCCTGGG 308
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 452 yLysLeuSerValAsnGlnLeu-GluThrSerAsnValAspMetSerArgGluMetVala 472
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 TTCGATCGTGGCCAACTCCCTNGGAGAACTCCAACGTCACCTGACCAAGAGTGTGG 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 472 snMetIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrAlaAspT 492
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 ACCTGATCAGGCGCCAGCAACTATCAGCGCAAGCGCAAGACCATCTCCACCCAAAGCA 428
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 492 hrMetLeuGlnLysAlaLeuLuleu 500
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 CCATCATGCAGACCATCATTCAGATG 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 18

```

CN582719/c
LOCUS

```

```

DEFINITION CN582719 661 bp mRNA linear EST 04-MAY-2004
USDA-PP 125782 Acyrthosiphon pisum, Pea Aphid Acyrthosiphon pisum
cDNA clone WHAP021_G07, mRNA sequence.

```

```

ACCESSION CN582719
VERSION CN582719.1 GI:46994442
KEYWORDS EST.
SOURCE Acyrthosiphon pisum (pea aphid)
ORGANISM Acyrthosiphon pisum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

```

```

REFERENCE 1 (bases 1 to 661)
AUTHORS Hunter,W.B., Dang,P.M., Stern,D. and Lapointe,S.L.
TITLE Gene expression in Acyrthosiphon pisum (Hemiptera: Aphididae)
JOURNAL Unpublished (2004)
COMMENT Contact: Wayne Hunter; Phat Dang
USDA-ARS
US Horticultural Research Lab, Ft. Pierce, FL
2001 South Rock Rd., Ft. Pierce, FL 34945, USA
Tel: (772) 462-5898, (772) 462-5940
Fax: (772) 462-5986
Email: Whunter@ushrl.ars.usda.gov.
Location/Qualifiers
1..661
/organism="Acyrthosiphon pisum"

```

```
/mol_type="mRNA"
/db_xref="taxon:7029"
/clone="WHA021_G07"
/sex="Mixed population"
/tissue_type="whole body, nymphs and adults"
/dev_stage="All"
/lab_host="Xli-Blue"
/clone_lib="Acyrthosiphon pisum, Pea Aphid"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; Acyrthosiphon pisum colonies. Library by Srin
Kambampati; Peter Dearden; David Stern, Department of
Ecology and Evolutionary Biology, Princeton University,
NJ. (609) 258-0759. A high quality EST with at least 100
contiguous bases at Trace Turner score of 20 or better.
Sequencing of clones by Dr. PM Dang, US. Horticultural
Research Lab, Ft. Pierce, FL. (772) 462- 5940."
```

ORIGIN

Alignment Scores:

Pred. No.:	1.6e-08	Length:	661
Score:	181.50	Matches:	58
Percent Similarity:	43.35%	Conservative:	30
Best Local Similarity:	28.57%	Mismatches:	88
Query Match:	7.02%	Indels:	27
DB:	7	Gaps:	6

US-10-009-823A-1 (1-502) x CNS82719 (1-661)

QY 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23

Db 599 TCAATATGATAGTACTAGTCTAGTCTAGCAATAATGATACATGGAATATATCC 540

QY 24 AsnAsnIleAlaAsnAlaThrIleGlyTyrGlyGlnGlnValValPheGlnAsp 43

Db 539 AATAATATTCCTAACGCATCAACTAGGATATAAATCTCGTAAACCTCTTTTTTTGAT 480

QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63

Db 479 ATGTTTCTCATCTATTTATCAATACACTACTAAAT-----GGA 441

QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 93

Db 440 TACGAGTGGGTATTTCAAGTATTATATACAAACTTTTAATAATGCGCATGTTAGTTGAAACT 381

QY 84 AsnSerValThrAspLeuAlaIleGlyGlyGlyPheGln---ValThrLeuGlu 102

Db 380 GGACGAGATTGGATTAGGAATTATAAAAGACGGCTTTTTCGCTCTTGAGACGTCAA 321

QY 103 AspLysValHisTyrTrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsn 122

Db 320 GGTATGTTCTATTATACAGAGATGGCAATTCTTCCTCGATAAGATCAAAATATTATC 261

QY 123 AspProSerGlyPheThrLeuMetGlySerArgIleSer-----Asn 136

Db 260 AATATTCAAGGTATGATCTAACTGGACTTAATACATCTTGTTCAAAAAGTGAATTTAAT 201

QY 137 AsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThr 156

Db 200 AAT-----AGATCCAACTTTAGAACCTTATTAATTA-----AAAAATCTAAT 159

QY 157 ValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGly--- 175

Db 158 ATTTTAAAAACAAACCTACTCTCTGAAATCATGTTTAAAGCGTTTGTGAAATCGTAATAC 99

QY 176 -----AspSerThrAspLysThrGlnSerGluAlaAsnProTyrPhe 189

Db 98 GAATCAAAACAGCTGGTGAATTTCTGACAAACAACTATCTATCAACCAAGATTACATG 39

QY 190 AlaLeuLeu 192

Db 38 ACTTATATT 30

RESULT 19

BP562788/c
LOCUS BP562788 RAFL14 Arabidopsis thaliana cDNA clone RAFL14-76-G17 5',
DEFINITION mRNA sequence.
ACCESSION BP562788
VERSION BP562788.1 GI:48978554
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 566)
AUTHORS Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
Nakajima, M., Enju, A., Akiyama, K., Ono, Y., Muramatsu, M.,
Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K.
FUNCTIONAL annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
JOURNAL 21932300
MEDLINE 11910074
PUBMED
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekior@c.riken.go.jp
Please visit our web site (http://pfgweb.gsc.riken.go.jp/) for
further details.
FEATURES
source location/Qualifiers
1. 566
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL14-76-G17"
/tissue_type="root"
/lab_host="DH10B"
/clone_lib="RAFL14"
/note="Site_1: BamHI; Site_2: SalI"

Alignment Scores:

Pred. No.:	2.95e-08	Length:	566
Score:	178.00	Matches:	36
Percent Similarity:	62.65%	Conservative:	16
Best Local Similarity:	43.37%	Mismatches:	31
Query Match:	6.88%	Indels:	0
DB:	5	Gaps:	0

US-10-009-823A-1 (1-502) x BP562788 (1-566)

QY 418 PheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAsp 437

Db 510 TTTATGATGACACCGGCTGGAGAGCATTTGGGAAACCTCTACACGAAACGCAATCC 451

QY 438 SerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLeuSerValAsn 457

Db 450 TGTGTGTCACCAAGCAAGCAACGCGGCTGTGAACGGCGGAGCTGCTGTATCAAGG 391

QY 458 GlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGln 477

Db 390 TATGTTGAACCTCTTAACGTCACGTCACGTCGCGGAGAACTGGTCAATATGATTCAGTGCA 331

QY 478 ArgGlyPheGlnMetAsnSerLysSerValThrAlaAspThrMetLeuGlnLysAla 497

Db 330 CGCGCTTACGAATCAACAGTAAAGCGGTGTCTCCACCAGCATCAGATGCTGCAAAACTG 271

QY 498 LeuGluLeu 500

Db 270 ACGCAACTC 262

RESULT 20

CD119056
LOCUS
DEFINITION
ME1-0052U-V142-G10-U.B. mRNA linear EST 14-SEP-2003
ME1-0052U-V142-G10.B. mRNA sequence.
CD119056
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Schistosoma mansoni
Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 436)
AUTHORS
Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acoelomate human parasite Schistosoma
mansoni
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjoe@iq.usp.br
This sequence was derived from the PAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: ME1-0052U-V142 row: 10 column: G.
FEATURES
Location/Qualifiers
source
1..436
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="ME1-0052U-V142-G10.B"
/sex="mixed pool"
/dev_stage="egg"
/lab_host="Mus musculus"
/clone_lib="ME1-0052"
/note="vector: pGEM T-easy"
ORIGIN
Alignment Scores:
Pred. No.: 4,08e-08 Length: 436
Score: 175.00 Matches: 33
Percent Similarity: 62.92% Conservative: 23
Best Local Similarity: 37.08% Mismatches: 33
Query Match: 6.77% Indels: 0
DB: 6 Gaps: 0
US-10-009-823A-1 (1-502) x CD119056 (1-436)
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyr 432
Db 2 GTTCGTGTGACACCTTTTGATATGCCAGGGCTGCAGCCCAATGGCGAAAACAACCTGG 61
QY 433 SerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGly 452
Db 62 GCATCTACTCCCTTCGGGTGGAGGCAAGCCCTGGCTCCCGGAGAGGAGGATTTTGGC 121
QY 453 LysLeuSerValAsnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAsn 472
Db 122 AAGTTGCAATCGTCTGCCATTGAAGAATCCACCGCATACCAACGAGAGGAGTGTATCAAC 181

QY 473 MetIleIleIleGlnArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db 182 ATGATGTCGCCAGCGCTTCTATCAGCCACGGCGAGACCATCAAGCAGGATGCC 241
QY 493 MetLeuGlnLysAlaLeuGluLeuLys 501
Db 242 ATCCTGCAAAACCTGCTCAACCTGGCT 268
RESULT 21
LOCUS
DEFINITION
ME1-0052U-V172-E05-U.B. mRNA linear EST 14-SEP-2003
ME1-0052U-V172-E05.B. mRNA sequence.
CD119187
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Schistosoma mansoni
Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 436)
AUTHORS
Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acoelomate human parasite Schistosoma
mansoni
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjoe@iq.usp.br
This sequence was derived from the PAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: ME1-0052U-V172 row: 5 column: E.
FEATURES
Location/Qualifiers
source
1..436
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="ME1-0052U-V172-E05.B"
/sex="mixed pool"
/dev_stage="egg"
/lab_host="Mus musculus"
/clone_lib="ME1-0052"
/note="vector: pGEM T-easy"
ORIGIN
Alignment Scores:
Pred. No.: 4,08e-08 Length: 436
Score: 175.00 Matches: 33
Percent Similarity: 62.92% Conservative: 23
Best Local Similarity: 37.08% Mismatches: 33
Query Match: 6.77% Indels: 0
DB: 6 Gaps: 0
US-10-009-823A-1 (1-502) x CD119187 (1-436)
QY 413 IleProLeuAlaArgPheThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyr 432

```

Db      2 GTTCGTGTACACCTTTGTAATGAGCCAGCGGCTGCAGCCCAATGCGGAAACAACTGG 61
Qy      433 SerAlaThrLeuAspSerGlyProGluPheGlyLeuProGlyThrSerAsnThrGly 452
Db      62 GCATCTACTCGCGTTCGGGTGAGCCCAAGCTGGCTTCGCGGAGAGAGCGAGTTTGGC 121
Qy      453 LysLeuSerValAnGlnLeuGluThrSerAsnValAspMetSerArgGluMetValAn 472
Db      122 AAGTTGCAATCGTTCGCAATGGAATCCAAAGCCGATACACCGGAGAGCTGATCAAC 181
Qy      473 MetIleLeuLeuArgGlyPheGlnMetAsnSerLysSerValThrThrAlaAspThr 492
Db      182 ATGATGTCGCGAGCGCTTCTATCAGCCCAAGCGGAGAGCATCAAGAGCGAGGATGCC 241
Qy      493 MetLeuGlnLysAlaLeuGluLeuLys 501
Db      242 ATCTGCAACCCCTGCTCAACCTCGT 268

RESULT 22
AI057966/c
LOCUS      532 bp mRNA linear EST 20-JUL-1998
DEFINITION SWOV13CAN21B03SK Onchocerca volvulus infective larva cDNA
            (SW94WL-OVL3) Onchocerca volvulus cDNA clone SWOV13CAN21B03 5',
            mRNA sequence.
ACCESSION  AI057966
VERSION     AI057966.1 GI:3331832
KEYWORDS   EST.
SOURCE      Onchocerca volvulus
            Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Onchocerca.
REFERENCE   1 Williams,S.A., Lu,W., Lizotte-Waniewski,M. and Laney,S.J.
            Genes expressed in infective third stage larvae of Onchocerca
            volvulus
JOURNAL     Unpublished (1995)
COMMENT     Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK
            Location/Qualifiers
                1..532
                    /organism="Onchocerca volvulus"
                    /mol_type="mRNA"
                    /strain="Giertra Leone"
                    /db_xref="taxon:6282"
                    /clone="SWOV13CAN21B03"
                    /lab_host="XL1-Blue MRP"
                    /clone_lib="Onchocerca volvulus infective larva cDNA
                    (SW94WL-OVL3)"
                    /note="vector: lambda Unizap XR; Site 1: EcoR I; Site 2:
                    Xho I; Cutaneous filarial nematode parasite of humans.
                    mRNA was prepared from third stage infective larvae of
                    Onchocerca volvulus isolated from mosquitoes 10 days after
                    infection and converted to double stranded cDNA using
                    reverse transcriptase and oligo(dT) followed by RNase H
                    and DNase I. The library had 1.8 x 108 independent
                    recombinants and average insert size was 900 base pairs.
                    The library was constructed by Wenhong Lu. The library is
                    available from Dr. S.A. Williams, email genome@smith.edu."

ORIGIN
Alignment Scores:
Pred. No.:      1.03e-07      Length:      532
Score:          172.50      Matches:      64
Percent Similarity: 39.57%      Conservative: 29
Best Local Similarity: 27.23%      Mismatches: 73

```

```

Query Match: 6.67%      Indels: 70
DB:          1          Gaps: 8

US-10-009-823A-1 (1-502) x AI057966 (1-532)

Qy      216 ValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSer 235
Db      516 GTTTCGACAGTACAGGTATGTCATGACATGAGCGTCTACTTT-----GTG 469
Qy      236 SerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGlySer 255
Db      468 AAGACCCGGGATATAACTGCGCAG-----GTCTACACCCAGGATAGCAGTGTCCAAAC 415
Qy      256 AlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSerSer 275
Db      414 AGCATTCGGAAGACA-----GCGACACACANTGGAAATTAAACGCT 376
Qy      276 AsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLysAspLeu 295
Db      375 AATGCGCACATTA-----GTGGATGGTGGCATGGCGAATAATATCCAAACCGCGCAATT 322
Qy      296 AsnAlaThrGlnProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheVal 315
Db      321 AACGGTCCAGAACCCGCC-----ACGTTTACTGTGAGCTTCCTC 283
Qy      316 GlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGlnAsnMetTyr 335
Db      282 AACTCC-----ATGCAGCAAAAT-----265
Qy      336 AlaGlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMet 355
Db      265 -----265
Qy      356 MetProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerThrArg 375
Db      264 -----ACCGCGCGTAAACAATATTGTGGCAACC-----238
Qy      376 ArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlu 395
Db      237 -----ACCAGAACGAGNAC-AAACCGCGGAGTCTGGTGGATTATCAATCAATGATGAC 185
Qy      396 GlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeu 415
Db      184 GGTACGGTTGTGCGCACTATTCCAAACGACAAACCAACTGCTGGGGCAGATTGTACTG 125
Qy      416 AlaArgPheThrSerGluAspGlyLeuArgArgGluGlyAsnAsnHisTyrSerAlaThr 435
Db      124 CGGAACNTTGCACACAGAGGTCTGGCATCCGAAAGGGAGACAACTGCTGGTCTGGCAG 65
Qy      436 LeuAspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsn 450
Db      64 CAATCTTCTGCGGTGGCGCTGTTGGGACAGCCCGGTGCACGTCA 20

RESULT 23
CD444654
LOCUS      715 bp mRNA linear EST 03-JUN-2003
DEFINITION ELO10N0442D08.b Endosperm_4 Zea mays cDNA, mRNA sequence.
ACCESSION  CD444654
VERSION     CD444654.1 GI:31360297
KEYWORDS   EST.
SOURCE      Zea mays
            Zea mays
            ORGANISM
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 715)
            Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,B. and
            Messing,J.
            Sequencing of the maize endosperm ESTs
            Unpublished (2002)
            Contact: Lai, Jinsheng
            Dr. Joachim Messing's lab
            Waksman Institute, Rutgers University

```

Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/cf.html>

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	

ORIGIN

Alignment Scores:	
Pred. No.:	3.85e-07
Score:	169.00
Percent Similarity:	64.00%
Best Local Similarity:	45.3%
Query Match:	6.54%
DE:	6
Gaps:	0
Indels:	0
Mismatches:	17
Conservative:	14
Matches:	34
Length:	715

US-10-009-823A-1 (1-502) x CD444654 (1-715)

427	GluclyAsnAsnHisTyrSerAlaThrLeuAspSerGlycylProGluPheGlyLeuPro	446
7	GAAGTGACACACGCTGCTGGACGCAATCTTCTGGCGTGTGGGACAGCC	66
447	GlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValAspMet	466
67	GGGACGGGAAACTTTGGACCCCTGACCAACGGTGGCTGGAAACGGTCCACATCGATCTC	126
467	SerArgGluMetValAsnMetIleIleGlnArgGlyPheGlnMetAsnSerLysSer	486
127	AGTAAAGAACTGGTGCAATATGATCGTGGCCAGCGTAACATACGTCTTAACGGCGGGACC	186
487	ValThrThrAlaAspThrMetLeuGlnLysAlaLeuGluLeuLys	501
187	ATCAAAACCCAGCAGCAGATCTTCACACAGCGTGGTTAACTTACCG	231

RESULT 24

Accession	Gene	Size (bp)	Library	EST ID
AF86932/c	LOCUS	692	mRNA	EST 03-NOV-2003
	DEFINITION	UI-CF-DUI-aas-f-05-18-UI.r18	UI-CF-DUI Homo sapiens	cdna clone
		UI-CF-DUI-aas-f-05-18-UI.5'	mRNA sequence.	

ACCESSION
CF886932
VERSION
CF886932.1 GI:38148793

KEYWORDS

SOURCE	ORGANISM
Homo sapiens (human)	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	1 (bases 1 to 692)
Bonaldi, M.F., Lennon, G. and Soares M.B.	Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)	JOURNAL

MEDLINE 97044477

PUBMED
8889548
Contact: McCray, PB
COMMENT

McCray Lab

University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

RESULT 25

BZ557676

LOCUS

DEFINITION

1
 2
 3
 4
 5
 6
 7
 8
 9
 10

ACCESSION
INTERPRETATION

VERSION
KEYWORDS

SOURCE

ORGANISM


```

ACCESSION      CN822868
VERSION        CN822868.1  GI:47950937
KEYWORDS       EST.
SOURCE         Ovis aries (sheep)
ORGANISM       Ovis aries
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
JOURNAL        Caprinae; Ovis; 1 (bases 1 to 747)
COMMENT        Gossner, A. and Hopkins, J.
                Ovine spleen/brain cDNA library
                Unpublished (2004)
                Contact: J Hopkins
                Veterinary Biomedical Sciences
                University of Edinburgh
                Summerhall Square, Edinburgh, EH9 1QH.
                Email: j.hopkins@ed.ac.uk
                Plate: 02 row: E column: 20
                Seq primer: M13reverse
                High quality sequence start: 7
                High quality sequence stop: 549.
FEATURES       Location/Qualifiers
                1..747
                /organism="Ovis aries"
                /mol_type="cDNA"
                /db_xref="taxon:9940"
                /clone="Oa脾bn.02E20"
                /clone_lib="Sheep spleen/brain pSport1 library"
ORIGIN
Alignment Scores:
Pred. No.:      1,628-05      Length:      747
Score:          154.00      Matches:    39
Percent Similarity: 46.43%      Conservative: 13
Best Local Similarity: 34.82%      Mismatches: 54
Query Match:     5.96%      Indels:     6
DB:              7          Gaps:         2
US-10-009-823A-1 (1-502) x CN822868 (1-747)
QY      4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
DB      426 TCTTTTAACTGGCGCTTATGGCGCTTATCGGCCCAACAACTGGCGTGCACCGGC 485
QY      24 AenAenIleAlaAsnAlaAsnThrIleGlyTyrlsGlnGlnValPheGlnAsp 43
DB      486 AACACATCGCCACGTCGCGACCAACCGCTTCAGTCGTCGCGATTCCTCGAT 545
QY      44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
DB      546 ATCTACGGCG-----CGCTCCAAAGCTGGCGCCCGCCAGAACACGATCGGC 590
QY      64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlyAlaPheGluProGly 83
DB      591 AACGCGCTAAACCTGGCGGAGTGTCCAGCAGTTCACCGAGGTGACGTCAACACAGC 650
QY      84 AenSerValThrAspLeuAlaIleGlyGlyLysGlyPhePhe---GlnValThrLeuGlu 102
DB      651 GCGCGCACGCTGGACATCGCATCGCATCGCGCGCGCTTCGTCGACAGAGCGCGAC 710
QY      103 AspLysValHisTyThrArgAlaGlyAsnPheArg 114
DB      711 GGTTCGCTGGAGTACACCCGTAACGTCGCTTCCTCCG 746
RESULT 29
AK044947
LOCUS
DEFINITION
Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
enriched library, clone: B130015D15 product: nucleoporin homolog
[Rattus norvegicus], full insert sequence.
ACCESSION      AK044947
VERSION        AK044947.1  GI:26336942
KEYWORDS       HTC; CAP trapper.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
REFERENCE      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL        Carninci, P. and Hayashizaki, Y.
COMMENT        High-efficiency full-length cDNA cloning
                Meth. Enzymol. 303, 19-44 (1999)
                99279253
                10349636
                Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
                Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
                Normalization and subtraction of cap-trapper-selected cDNAs to
                prepare full-length cDNA libraries for rapid discovery of new genes
                Genome Res. 10 (10), 1617-1630 (2000)
                20499374
                11042159
                Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
                Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
                Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
                Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
                Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuura, M.,
                Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
                Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
                RIKEN integrated sequence analysis (RISA) system--384-format
                sequencing pipeline with 384 multicapillary sequencer
                Genome Res. 10 (11), 1757-1771 (2000)
                20530913
                11076861
                The RIKEN Genome Exploration Research Group Phase II Team and the
                FANTOM Consortium.
                Functional annotation of a full-length mouse cDNA collection
                Nature 409, 685-690 (2001)
                The FANTOM Consortium and the RIKEN Genome Exploration Research
                Group Phase I & II Team.
                Analysis of the mouse transcriptome based on functional annotation
                of 60,770 full-length cDNAs
                Nature 420, 563-573 (2002)
                6 (bases 1 to 3057)
                Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
                Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
                Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
                Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
                Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
                Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
                Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
                Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
                Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
                Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
                Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
                Muramatsu, M. and Hayashizaki, Y.
                Direct Submission
                Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
                Physical and Chemical Research (RIKEN), Laboratory for Genome
                Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
                RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
                Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
                URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
                Fax: 81-45-503-9216)
                cDNA library was prepared and sequenced in Mouse Genome
                Encyclopedia Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                Division of Experimental Animal Research in Riken contributed to
                prepare mouse tissues.
                Tissues were provided by Dr. Tomohiro Kono (Department of Animal
                Science, Tokyo University of Agriculture, 1737 Hunko Atsugi City,
                Kanagawa Prefecture, Japan) whose assistance we gratefully
                acknowledge.
                Please visit our web site for further details.

```

```

URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
  Source
    1..3057
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="FANTOM_DB:BI30015D15"
      /db_xref="taxon:10090"
      /clone="BI30015D15"
      /tissue_type="parthenogenote"
      /clone_lib="RIKEN full-length enriched mouse cDNA library"
      /dev_stage="9.5 days embryo"
      1..2002
      /note="unnamed protein product; nucleoporin homolog
      [Rattus norvegicus] [PIR|A44345, evidence: PASTY, 92.8%ID,
      45.4%length, match=1995]
      putative"
      /codon_start=2
      /protein_id="BAC32155.1"
      /db_xref="GI:26335943"
      /translation="CNKAEDNRVCSTSEKPLVSAASSGPAPVLSGGCLGLDKFK
      KPEGSWDEVCLVQNKADSAKCIACSAKPGTKSEFKGFTGSSSLNPAPSAFKFGIPS
      SSGSLQTLSTGNFKFGOGGLTSSDGSNTMTNFKSKKEELPKSPAGFSPFA
      KPEEVKNDKNDTFQSGSLTNPASSAPQFQSVTLGOQKEELPKSPAGFSPFA
      AGVNNPNAADITATSENKSGFNGLDTSKVSVPFTVKTTEAKKEDAPATKGF
      FGKVGSSLPSSMFLVGRTEKQOEPVTSLSVFGKADSEEPKQPVSPFNSSQGT
      KDESKPFPSVAKPGSEQLQAKATFAPGNQNTITDQGAAPVPSFLNSSSSS
      SPATSSGGPIFGSTSSNPVAAFFVQGNAPVSSAFGNAESSTQSLLFPOES
      EPATSPAPAAFPVFTGASSNSVSSGFTFGATTTSSSGSFFVGTGHSAPASAP
      AFGANQPTFQSGQASQNPFPFGSISSTALFSAQSQVPPPIFGTVSSSQPPVF
      QOQPSQAFSGGTANASVFOFGSTTNFNFTNNPSGVFTFGASPTPAPASQPSGS
      GVFSQSOSPASTVFGSGNKMFFSSSGTSVSGRKLKTAVERKK"
      3039..3044
      /note="putative"
      3057
      /note="putative"
      ORIGIN
      Alignment Scores:
      Pred. No.:      Length:      3057
      Score:          Matches:      122
      Percent Similarity: 31.66%
      Best Local Similarity: 22.72%
      Query Match:      5.38%
      DB:               Indels:      150
                        Gaps:        23
      US-10-009-823A-1 (1-502) x AK044947 (1-3057)
      QY      22 ValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLysGlnGlnValValPhe 41
      Db      494 GTTAAACACGACACAAAGATGATAAT-----TTT 523
      QY      42 GlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGln 61
      Db      524 CAG-----TTGGATCTCTTCGGCTTAACCAACCCAGCTTCTCAGCTCCGTTTCAG 577
      QY      62 AlaGlyMetGlyAla-----GlnValGlySer 70
      Db      578 TTTGGGTATCTACTCTTGGCAACAAGAAAGAGAGGAGCTGCCTAAATCTTCACT 637
      QY      71 ValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAla 90
      Db      638 GCAGGCTTTAGCTTGTCTGCGAGGTGTTAATAACCCCTTAATGCTGCTATTGACACCACA 697
      QY      91 IleGlyGlyLys-----GlyPheGlnValThrLeuGluAspLys---ValHis 106
      Db      698 GCGACCTCAGAGAACAGACGCGCTTCACTTTGGAAACCCCTAGACACAAAGAGTGCTCA 757
      QY      107 TyrThr-----Argala 110
      Db      758 GTGACTCTTTTCATACAGACACAGAGCAAGAAAGAAAGACGACGACCTGCTACCAAG 817
      QY      111 GlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMet 130
      Db      818 GGTGGGTTCACATTGTGTAAGTGGGCTCTCTCTTCTTGCCATCTTCTCAATGTTGT 877
      QY      131 GlySerArg----- 133
      Db      878 TTGGGAAGGACAGAGAGGAGAAACAGCAAGAACCTGTGTACTTCTCTGTGTGTTGGG 937
      QY      134 -----IleSerAsnAsnPro 138
      Db      938 AAGAAAGCCGACAGTCAGGAGCAAGAGTCTCAGCCAGTATTTTCTTTGGGAATTCAGAG 997
      QY      139 AsnIleGlySerGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAla 158
      Db      998 CAAACCAAGATGAGAGCTCCAGCGG---ACATTTCAGCTC-----AGTGTGGCA 1045
      QY      159 LysSerProAlaIlySerThrAlaLeu---AsnAlaValValAsnLeuGlyAspSer 177
      Db      1046 AAACCACTCTGGGAAGAGTCTCAGCAGCTAGCCAGGCTACTTTTGGCATTTGGAATCAA 1105
      QY      178 ThrAspLysThr-----GlnSerGluAlaAsnProTyrPheAlaLeuGluSerTyr 195
      Db      1106 ACCAATACAACTGATCAAGGTGCAGCAAGCCAGTTTTTAGCTTCTTGAACAGCAGT 1165
      QY      196 LysGlyAsnGlyThrProProIle-----SerThr 205
      Db      1166 TCCTCTAGTCAAGTGCACGACCACTTCACTAGTGTGGCATATTTGGTAGTTCAC 1225
      QY      206 SerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisAsp 225
      Db      1226 TCTTCTCCAAACCCACTGTGCACCTTTGTGTTGGACAGCCAGCAATCTGTCTCAGC 1285
      QY      226 IleThrValTyrPheAspGlyAlaProSerSerThrGly----- 238
      Db      1286 AGCTCTGCCTTCGGTAACGCTGCTGAATCCAGTACATCTCAGTCTTGTCTATTTCTCAA 1345
      QY      239 -----SerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
      Db      1346 GAGAGCGAGCCAGCCACCAGCTCCAGCAGCTGCCCGAGCTCTCTCTTGTATTGTTGG 1405
      QY      255 SerAlaAlaSerGlyThrAsp---SerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
      Db      1406 ACAGAGCGACGAGTAAATAGTGTCTCTCTGGGTCTTCTTGGGGTCTACAAACCGTCA 1465
      QY      274 SerSerAsnGlyGluLeu-----LysAsnMetThr 283
      Db      1466 AGCTCTTCAGGATCTCTCTTGTATTTGGCACTGGACATTCAGCACCATTCGCAGCTCCA 1525
      QY      284 AlaPhe-----ThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGln 299
      Db      1526 GCATTTGGTGTAAACAGACTCCAACTTTGGCAAAAGTCAA-----GGTGCCAGCCAG 1579
      QY      300 ProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyLe 319
      Db      1580 CCTAACCCGCAAGCTTGGCTCTATCTGCTTCAACAGCATTTATTTCTGTGTGTCT 1639
      QY      320 GlnProLeuThrLeuAsp---PheGlyLe-----LysSerGlnGlnAsnMetTyrAla 336
      Db      1640 CAGCTGTACCACCACTATATTTGGGACAGTGTCAAGCAGCAGCCACCTCTCTGTGTTT 1699
      QY      337 GlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMet 356
      Db      1700 GCACAGCAGCCCTAGCTAGCTGCTGCTTTGGCTCTGGG----- 1735
      QY      357 ProIleGlnThrSerSerGlyAsnSerThrAlaAsnGlySerSerSerThrArgArg 376
      Db      1736 -----ACAGTAAAGCAGTTCGGTTTCCATTTGGTAGCAGCATAGAAATTC 1786
      QY      377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGly 396
      Db      1787 AACTTCAAAACATAATCCATCAGGAGTGTTCACATTTGCTGCAAGTCCAGCAGCGCT 1846
      QY      397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla 416

```

```

Db      1847 GCAGCTCAGCCAGCCCTCAGGTCAGGGGTCTTTTCATTACGCCAGTCTCCA---GCA 1903
QY      417 ArgpThrSerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeu 436
Db      1904 TCATTTTACA-----GTGGGGTCAAAATGGGAAATATGTTCTCTCTTC--- 1948
QY      437 AspSerGlyGlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLys 453
Db      1949 -----GGAACTTCAGTTCTGTGGTCGC 1969

RESULT 30
LOCUS   AZ577776
DEFINITION 14f11 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
ACCESSION AZ577776
VERSION   1
KEYWORDS  374 bp DNA linear GSS 08-DEC-2000
SOURCE   sp. NGR234 genomic clone 14f11, genomic survey sequence.
ORGANISM Rhizobium sp. NGR234
          Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
          Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
REFERENCE 1 (bases 1 to 374)
AUTHORS  Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
TITLE    Genetic snapshots of the Rhizobium species NGR234 genome
JOURNAL  Genome Biol. 1 (6), RESEARCH0014 (2000)
MEDLINE  21114532
PUBMED  11178268
COMMENT  Contact: Virginie Viprey
          Laboratoire de Biologie Molculaire des Plantes Supérieures
          University of Geneva
          1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
          Tel: +44(0)1603450000
          Fax: +44(0)1603450045
          Email: virginie.viprey@bsrc.ac.uk
          Class: shotgun.
FEATURES
         source
         location/Qualifiers
             1..374
             /organism="Rhizobium sp. NGR234"
             /mol_type="genomic DNA"
             /strain="ANU265"
             /db_xref="taxon:394"
             /clone="14f11"
             /clone_lib="Shot-gun genomic library of Rhizobium strain
             ANU265"
             /note="Vector: ML3; derivative strain of NGR234 cured of
             pNGR234a"
ORIGIN
Alignment Scores:
Pred. No.: 0.000275 Length: 374
Score: 138.00 Matches: 28
Percent Similarity: 56.76% Conservative: 14
Best Local Similarity: 37.84% Mismatches: 32
Query Match: 5.34% Indels: 0
DB: 8 Gaps: 0

US-10-009-823A-1 (1-502) x AZ577776 (1-374)
QY      420 SerGluAspGlyLeuArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGly 439
Db      5 AACAGGCGCGGTGGAGCCGCGGCGGAGAACTCTTCAGCAGACAGCGCATCGGC 64
QY      440 GlyProGluPheGlyLeuProGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeu 459
Db      65 GAACCGATCGTCGGCAGCGCGCGCGATCCCGCTCCGCGCAGATCAAGCAGGCGTATCTG 124
QY      460 GluThrSerAsnValAspMetSerArgGluMetValAsnMetIleIleGlnArgGly 479
Db      125 GAATCTTCAAGTCAAGTCCGTCAGGAAATACCCGACCTGATCTCGGCCGCGCGCC 184
QY      480 PheGlnMetAsnSerLysSerValThrThrAlaAspThrMet 493

```

```

Db      195 TACGAATGAATCAAGTCTCATCCAGCGCGCGAGATG 226
QY      10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db      169 ACTGG---AAGGTGACTCCAGCGGCTTCAGTAGCTACAGTGGCTCCAGCAGTTCTGGC 225
QY      30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db      226 AGCTCCATTTCCAGTCCAGAGC----- 249
QY      50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db      250 -----TCTGGTGGTGGCTCCAGTGGTAGCTCCAGCGGATCCAGCATTCGCCAGGTGT 303
QY      70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db      304 TCTGCA-----GGATCTTTTAAGCCAGGAAACGGGGTATTCACAGGTC 345
QY      90 AlaIle-----GlyGlyLysGlyPheGlnVal--- 99
Db      346 AGCTACTCTCCGATCTGGCTCTAGTCTACAAAGTGATCCCGCTTCTCTCCAGCTGGGG 405

RESULT 31
LOCUS   AY412935
DEFINITION Homo sapiens CDSN gene, VIRTUAL TRANSCRIPT, partial sequence,
ACCESSION AY412935
VERSION   1
KEYWORDS  1590 bp DNA linear GSS 16-DEC-2003
SOURCE   genomic survey sequence.
ORGANISM Homo sapiens (human)
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1590)
AUTHORS  Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
          Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
          Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
          Adams,M.D. and Cargill,M.
TITLE    Inferring nonneutral evolution from human-chimp-mouse orthologous
          gene trios
JOURNAL  Science 302 (5652), 1960-1963 (2003)
PUBMED  14671302
REFERENCE 2 (bases 1 to 1590)
AUTHORS  Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
          Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
          Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
          Adams,M.D. and Cargill,M.
TITLE    Direct Substitution
JOURNAL  Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
          Rockville, MD 20850, USA
COMMENT  This sequence was made by sequencing genomic exons and ordering
          them based on alignment.
FEATURES
         source
         location/Qualifiers
             1..1590
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             <1..>1590
             /gene="CDSN"
             /locus_tag="HCM4723"
ORIGIN
Alignment Scores:
Pred. No.: 0.003 Length: 1590
Score: 137.50 Matches: 113
Percent Similarity: 32.64% Conservative: 59
Best Local Similarity: 21.44% Mismatches: 185
Query Match: 5.32% Indels: 171
DB: 9 Gaps: 24

US-10-009-823A-1 (1-502) x AY412935 (1-1590)
QY      10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db      169 ACTGG---AAGGTGACTCCAGCGGCTTCAGTAGCTACAGTGGCTCCAGCAGTTCTGGC 225
QY      30 AsnThrIleGlyTyrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db      226 AGCTCCATTTCCAGTCCAGAGC----- 249
QY      50 AlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db      250 -----TCTGGTGGTGGCTCCAGTGGTAGCTCCAGCGGATCCAGCATTCGCCAGGTGT 303
QY      70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db      304 TCTGCA-----GGATCTTTTAAGCCAGGAAACGGGGTATTCACAGGTC 345
QY      90 AlaIle-----GlyGlyLysGlyPheGlnVal--- 99
Db      346 AGCTACTCTCCGATCTGGCTCTAGTCTACAAAGTGATCCCGCTTCTCTCCAGCTGGGG 405

```

QY 100 -----ThrLeuAlaAspLysValHisTyr 107
 Db 406 AGCAGCAGCTCTCACTCGGGAAGCAGCGCTCTCACTCGGGAAGCAGCAGCTCTCATTCG 465
 QY 108 ThrArgAlaGlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPhe 127
 Db 466 AGCAGCAGCAGCAGCTTTTCAGTTTCAGTCAGCAGCAGCAGCTTCCAA----- 507
 QY 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysLysLeuThrLeuGluPro 147
 Db 508 GTAGGGAATGGCTCTCTGTCGCAACCAATGACACTCTTACCGCGGATATAACACCT 567
 QY 148 IleGln-LeuAsp-----PheAsnAspProThrValAlaLys 159
 Db 568 TCCAGCCTGGACAAAGCTTTCCTCTTCCCAAACTTTTGGGTATCCAGCAGTGGCAA 627
 QY 159 sSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeuGlyAspSerThr-- 178
 Db 628 ACGTCAGCTCAACACGAGCGTCCCTGTA-----GTTCG-CACATCCCGACTCTCCCTG 680
 QY 179 -----AspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSerTrpLys 196
 Db 681 CAGTGGAGGGCCCAATCGCTCGCACTCTGCGCCCTTACATCCCAAGCTCCCACTCTGTGTC 740
 QY 196 sGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMetArgVa 216
 Db 741 AGGGGTCAGAGGCTGTGGTGGTG-----GT 767
 QY 216 lTyAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaProSerSe 236
 Db 768 GTTGACACGACGAGCTTCTGTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 821
 QY 236 rThrGly-----SerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAs 253
 Db 822 CAATGGTGGCTTCCAGGCAAGCCCTGTGCCCCCACTCACTGTATAGCAAACTCTATGG 881
 QY 253 pGlySerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPh 273
 Db 882 TGGCTACGAGTGTGGTGGCTCTCTGACAGTTATCTGTTCCAGGC---ATGACCTA 938
 QY 273 eSerSerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLy 293
 Db 939 CAGTAAGGGT-----AAATCTATCTCTGCTGGCTTACTTCCACCA 977
 QY 293 sAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuProGlnPhe----- 310
 Db 978 AGAG--AACCTGTGAAGGCTCTCA-----GGGCTCTCTCTCTCTCTCTCTCT 1025
 QY 311 -----SerAl 312
 Db 1026 GCCCCCATCTCTGAGGCAAAATCTTCTCCAGCAACCCCATCATCCCCAGCAGTCGGC 1085
 QY 312 aAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIleLysSerGlnGl 332
 Db 1086 AGCTTCTCGGCAATTCGTTCCAGCAGTG-----GGGACTGTGGTGGGTCCA 1133
 QY 332 nAsnMetTrpAlaGlyAlaProAlaSerAlaAlaAlaIleGlyThrAspIleGlyLysLe 352
 Db 1134 GCTCTGTGGAGGCGGCTCCAGCGGCTCCAGGACCTGTCTCTCCCTCCAGTTCCTGAGT 1193
 QY 352 uProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaAsnGlySerSe 372
 Db 1194 CCCAGAGGTTTACGATTTCAGCATTTCCAGCAGCTCCGCTTCCACCTACCAATCCCTGGCGAGTGC 1253
 QY 372 rSerThrArgTyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrI 392
 Db 1254 TTCCAGAGCCCTGCTCCCAACCCAGGC----- 1281
 QY 392 eThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAs 412
 Db 1282 -----ACCGCTCTCTTCCAGCAGCTCC----- 1305

QY 412 nileProLeuAlaAArgPheThrSerGluAsp-----Gly 423
 Db 1306 -----AGTTCCAATCGAGTGGCAAAATCATCTTCAGCCTGTGG 1346
 QY 423 YLeuAArgGluGlyAsnAsnHisTyrSerAlaThrLeuAspSer-----Gly 439
 Db 1347 CAGCAAGTCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1406
 QY 439 YGlyProGluPheGlyLeuPro-----GlyTh 448
 Db 1407 GGGCCCCGAT--GGTCTCCCACTCTGATCCCTCGCTGGTGGCAAGCCCTGTGGCTC 1463
 QY 448 rSerAsnTyrGlyLysLeu 454
 Db 1464 CAGCAGTGTGGAAGATC 1482
 RESULT 32
 LOCUS AY416869
 DEFINITION Pan troglodytes NUP153 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
 ACCESSION AY416869
 VERSION AV416869.1 GI:39772829
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 REFERENCE 1 (bases 1 to 4428)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 4428)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
 FEATURES
 source
 locus_tag="HGM6034"
 gene
 locus_tag="HGM6034"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0244 Length: 4428
 Scores: 135.50 Matches: 112
 Percent Similarity: 34.15% Conservative: 56
 Best Local Similarity: 22.76% Mismatches: 221
 Query Match: 5.24% Indels: 103
 DB: 9 Gaps: 21
 US-10-009-823A-1 (1-502) x AY416869 (1-4428)
 QY 16 SerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAlaAsnThrIleGlyTyrLys 35
 Db 3067 AGCTTTGGTACAGGTGTATTAACTCCACCCCTGCTCTGCTTAACACCATAGTGACCTCT 3126
 QY 36 GlnGlnGlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThrGly 55

```
Db 3127 GAGAACAGAGCAGCTTCAAC-----TTTGGACACACAGAA 3162
Qy 56 SerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArg----- 72
Db 3163 ACCAAGAGTCCTTACGTGCTCTTTCACATGTAGACATCAGAAAGTAAAGAAAGAA 3222
Qy 73 ThrIlePheThrGlnGlyAlaPheGluProGlyIleAsnSerValThrAspLeuAlaIleGly 92
Db 3223 ATGCCTGCCACCAAGAGGAGTTCCTTTTGGCAAC---GTGGAGCCTGCTCTCTGCCA 3279
Qy 93 GlyIleGlyPhePheGlnVal-----ThrLeuGluAspIleValHisThrArgAla 110
Db 3280 TGTGCTCAGTGTGTTTGTGGAGGACAGAAAGCAACCAAGAGCCTGTCACTTCT 3339
Qy 111 GlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSer-----GlyPhe 127
Db 3340 ACTTCTCTAGTTTGTGGAGAAAGAGCTGACAATGAAGAGCCAAAGTGTCAACCAAGTGT 3399
Qy 128 ThrLeuMetGlySerArgIleSerAsnAsnProAsnIleLysIleGluThrLeuGluPro 147
Db 3400 TCTTTGGGAATTCAGACAAACAAAGATGAGATTTCTTCAAGTCCACA----- 3450
Qy 148 IleGlnLeuAspPheAsnAspProThrValAlaLysSer-----ProAlaLysThrSer 165
Db 3451 TTTAGTTTGTATGATCACAACCACTCTGAGAAGGAATCTGAACAGCCAGCAAAAGCCACT 3510
Qy 166 ThrAlaLeuAsnAlaValAlaAsnLeuGlyAspSerThrAspIleThrGlnSerGluAla 185
Db 3511 TTTGCTTTGGAGCTCAAACT-----AGTACTACAGCTGATCAAGGTGCAGCA 3558
Qy 186 AsnProTyPheAlaLeuLeuGluSerTrpLysGlyAsnGlyThrProPheIle----- 203
Db 3559 AAGCCAGCTTTTGTGTTCTTGAACAACAGTTCCTCTAGTTCACAGTACACCACTCTCT 3618
Qy 204 -----SerThrSerAsnTySerTyAlaGlnProMetArg 215
Db 3619 GCTGTGTGGGATATTTGTAGTTCACCTCTTCTCCCAATCCACCTGTGGGTACCTTT 3678
Qy 216 ValTyAspGlnGlnGlyAsnSerHisAspIleThrValTyPheAspGlyAlaProSer 235
Db 3679 GTGTTGGACAGTCCAGCAATCCTGTGAGCAGCTCTGCTTGTGTAACACTGTGTAATCC 3738
Qy 236 SerThrGly-----SerLysThrPheGluTy 244
Db 3739 AGCACTCTCAGCTCTTCTGCTTTCTCAAGATAGCAAACTAGCAACCACTCCAGCA 3798
Qy 245 LeuValAlaMetAsnProSerGluAspGlySerAlaAlaSer-----GlyThrAspSer 262
Db 3799 GGTACAGCTGTCAACCCATTTGCTTTGTCTCCAGGAGCCAGCAGTATATACTACCACC 3858
Qy 263 AlaGlyLeuLeuMetSerGlyThrMetThrPheSerSerAsnGlyGluLeu----- 279
Db 3859 TCTGTTTGGGCTTTGGAGCCCAACCACTATCTAGCTCTGAGGATCCTCTTTGTATTT 3918
Qy 280 -----LysAsnMetThrAlaPhe-----ThrProThr 288
Db 3919 GGAAGTGGACCTCAGCACCATCTGCCAGTCCAGCATTTGTGTCTAACCAGACCCCAACA 3978
Qy 289 GlySerAlaThrLysAspLeuAsnAlaTrpGlnProAlaProLeuValAsnGlyLeuPro 308
Db 3979 TTTGGCAAAAGTCAA-----GTGTGCCAGCCAGCCCAATCCCCAGGCTTTGGATCTATA 4032
Qy 309 GlnPheSerAlaAsnPheValGlyAlaGlyIleGlnProLeuThrLeuAspPheGlyIle 328
Db 4033 TCATCTTCCACAGCATTTATTTCCCACTGTTCTCAGCTGCACCACTCTTTTGGACA 4092
Qy 329 LysSerGlnGlnAsnMetTrpAlaGlyAlaProAlaSerAlaAlaIleGlyThrAsp 348
Db 4093 GTGTCAAGCAGTACAGCAG-----CCCCCTGTG----- 4119
Qy 349 IleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerThrAlaArg 368
Db 4120 TTTGACAGCAACCTAGT-----CAGTCTGCATTTGGCTCTGGAACTCTCT 4167
```

```
Qy 369 Asn-----GlySerSerSerThrArgArgTySerGlnAspGly 381
Db 4168 AATTCTANNNNNNNTTCCAGTTTGGCAGCAGCACATACAAATTCNNCTTCAAAACAAC 4227
Qy 382 TyrProGlnGlyAspLeuValAspValThrIleThrSerGluGlyLysLeuGlnGlyLys 401
Db 4228 AATCCATCAGGAGTGTTCACATTTGGTGCAAAATTTAGCACACCTGCAGCCTCAGCCAG 4287
Qy 402 TyrSerAsnSerGlnValValAspPheTyAsnIleProLeuAlaArgPheThrSerGlu 421
Db 4288 CTTTCCAGCTCGGGGCTTTCATTTAACAGTCTCCA---GCAGCATTTACA----- 4338
Qy 422 AspGlyLeuArgGluGlyAsnAsnHisTySerAlaThrLeuAspSerGlyPro 441
Db 4339 -----GTGGGCTCAATGGMAAATGTCTTCTCTTCTCT----- 4374
Qy 442 GluPheGlyLeuProGlyThrSerAsnTyGlyLys 453
Db 4375 -----GGAATTCATTCTCTGGTCGC 4395

RESULT 33
CN823464          726 bp      mRNA      linear      EST 02-JUN-2004
LOCUS            Oa_splbn_05M02_M13reverse Sheep spleen\brain pSport1 library Ovis
DEFINITION      aries cDNA clone Oa_splbn_05M02 5', mRNA sequence.
ACCESSION       CN823464
VERSION         CN823464.1 GI:47951533
SOURCE          EST.
ORGANISM        Ovis aries (sheep)
REFERENCE       1 (bases 1 to 726)
AUTHORS         Gosner, A. and Hopkins, J.
TITLE           Ovine spleen\brain cDNA library
JOURNAL         Unpublished (2004)
COMMENT         Contact: J Hopkins
                Veterinary Biomedical Sciences
                University of Edinburgh
                Summerhall Square, Edinburgh, EH9 1QH.
                Email: j.hopkins@ed.ac.uk
                Plate: 05 row: M column: 02
                Seq primer: M13reverse
                High quality sequence start: 15
                High quality sequence stop: 547.
FEATURES        Location/Qualifiers
                 1..726
                 /organism="Ovis aries"
                 /mol_type="mRNA"
                 /db_xref="taxon:9940"
                 /clone="Oa_splbn_05M02"
                 /clone_lib="Sheep spleen\brain pSport1 library"

ORIGIN
Alignment Scores:
Pred. No.:      0.00234      Length:      726
Score:          133.50      Matches:    32
Percent Similarity: 45.74%      Conservative: 11
Best Local Similarity: 34.04%      Mismatches: 46
Query Match:    5.16%      Indels:     5
DB:              7          Gaps:      1

US-10-009-823A-1 (1-502) x CN823464 (1-726)

Qy 4 SerLeuPheIleGlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSer 23
Db 432 TCTTTTAACTCGGCTCTTAGCGGCTCTATGCGGCAACAACAACACTGACGTCGCGC 491
Qy 24 AsnAsnIleAlaSerAlaAsnThrIleGlyTyLysGlnGlnValValPheGlnAsp 43
Db 432 AACACATCGCACGTCGCGGACCAACCGGCTTCAATGTCGCGCGCAATTCGCGAT 551
```

```

QY 44 LeuPheSerGlnAspLeuAlaIleGlySerThrGlySerGlnGlyProAsnGlnAlaGly 63
Db 552 ATCTACGCG-----GGTCCAGCTGGGACCGCGCAGAACAGATCGCG 596
QY 64 MetGlyAlaGlnValGlySerValArgThrIlePheThrGlnGlnGlyAlaPheGluProGly 83
Db 597 AACGGCGTAAACCTGGCGGAGCTGTCCAGCAGTTTCAACCCAGGTTGACGTCAACACACAGC 656
QY 84 AsnSerValThrAspLeuAlaIleGlyGlyLysGlyPhePhe 97
Db 657 GCGGCGACCTGGACATGGCGATCCAGGCGCGGCTTCTTC 698

RESULT 34
LOCUS AY416870
DEFINITION Mus musculus NUP153 gene, VIRTUAL TRANSCRIPT, partial sequence,
ACCESSION AY416870
VERSION AY416870.1 GI:39772830
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..4355
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..4355
/gene="NUP153"
/locus_tag="HCM6034"

gene
ORIGIN
Alignment Scores:
Pred. No.: 0.0388 Length: 4355
Score: 133.50 Matches: 110
Percent Similarity: 33.75% Conservative: 53
Best Local Similarity: 22.77% Mismatches: 217
Query Match: 5.16% Indels: 103
DB: 9 Gaps: 19

US-10-009-823A-1 (1-502). x AY416870 (1-4355)
QY 16 SerThrGlyLeuGlyThrValSerAsnAlaIleAla-----AsnAlaAsnThrIle 32
Db 2890 TCCTCTGGCTTACCAACCCAGCTTCTTCAGCTCCGTTTCAGTTGGGGTATCTACTCTT 2949
QY 33 GlyTyrLysGlnGlnValPheGlnAspLeuPheSerGlnAspLeuAlaIleGly 52
Db 2950 GGGCAACAAGAAAGAGAGAGAGCTGCTAAATCTTCACTGCAGGCTTTAGCTTTGGT 3009
QY 53 SerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArg 72

```

```

Db 3010 GCAGGTGTTTAATACCCCTTAAT-----GCTGCTATTGACACACAGCG 3054
QY 73 ThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeuAlaIleGly 92
Db 3055 ACCTCAGAGAACAGAGCGCTTCACTTTGGAACCTAGACACAAG---AGTGTCTCA 3111
QY 93 GlyLysGlyPhePheGlnValThrLeuGluAspLysValHisTyrThrArgAla----- 110
Db 3112 GTGACTCCTTTTCATACAGACAAAGAGAAAGAAAGAAAGAGACACCTGCTCAAG 3171
QY 111 GlyAsnPheArgPheThrGlnAspGlyPheLeuAsnAspProSerGlyPheThrLeuMet 130
Db 3172 GGTGGCTTCACATTGGTAAAGTGGCTCTCTCTTCCCATCTTCCCAATGTTGT 3231
QY 131 GlySerArg----- 133
Db 3232 TTGGGAAGGACAGAGAGAGAAACAGCAAGAACCTGTACTTCTTCTCTGTTGGG 3291
QY 134 -----LleSerAsnAsnPro 138
Db 3292 AAGAAGCCGACAGTGCAGGAGCAAGTGTGCAGCAGTATTTCTTGGGAATTCAGAG 3351
QY 139 AsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAspProThrValAla 158
Db 3352 CAACCAAGATGAGAGCTCCAAAGCG--ACATTTCAGCTTC-----AGTGTGCA 3399
QY 159 LysSerProAlaLysThrSerThrAlaLeu---AsnAlaValValAsnLeuGlyAspSer 177
Db 3400 AAACCATCTGGAGAGAGTCTGAGCAGTACGCAAGGCTACTTTGCAATTTGGAAATCAA 3459
QY 178 ThrAspLysThr-----GlnSerGluAlaAsnProTyrPheAlaLeuGluSerTyr 195
Db 3460 ACCAATACAACTGATCAAGGTGCAGCAAGCCAGTTTITAGCTTCTTGAACAGCAGT 3519
QY 196 LysGlyAsnGlyThrProIle-----SerThr 205
Db 3520 TCCTCTAGTTCAAGTGCACCGCCACTTCATCTAGTGTGGCATATTTGTTAGTTCCACC 3579
QY 206 SerAsnTyrSerTyrAlaGlnProMetArgValTyrAspGlnGlnGlyAsnSerHisasp 225
Db 3580 TCTTCTCCACCCACCTGTGGCGCCCTTGTGTGGACAGCCAGCAATCTGTGTCAGC 3639
QY 226 IleThrValTyrPheAspGlyAlaProSerSerThrGly----- 238
Db 3640 AGCTCTGCTTCGGTAAAGCTGCTCAATCCAGTACATCTCAGTCTTGTGCTATTTCTCAA 3699
QY 239 -----SerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
Db 3700 GAGAGCGAGCCAGCCACACCTCCAGCAGTCCCGCCAGCTGCTCTCCCTTTGTTATTTGGG 3759
QY 255 SerAlaAlaSerGlyThrAsp---SerAlaGlyLeuLeuMetSerGlyThrMetThrPhe 273
Db 3760 ACAGAGCCAGCAGTAAATAGTGTCTCTCTGGGTTCACATTTGGGGCTACACACGTC 3819
QY 274 SerSerAsnGlyGluLeu-----LysAsnMetThr 283
Db 3820 AGCTCTTCAGATCTCTCTTTGTTATTTGGCAGTGCAGCAATTCAGCAATTCGACGTCCA 3879
QY 284 AlaPhe-----ThrProThrGlySerAlaThrLysAspLeuAsnAlaTyrGln 299
Db 3880 GCATTTGGTGTAAACAGACTCCACACATTTGGCAAAAGTCAA-----GGTGCAGCCAG 3933
QY 300 ProAlaProLeuValAsnGlyLeuProGlnPheSerAlaAsnPheValGlyAlaGlyIle 319
Db 3934 CCTAAACCGCAAGCTTTGGCTCTATCTCTCTTCAACAGCATTTATTTCTGCTGTTCT 3993
QY 320 GlnProLeuThrLeuAsp---PheGlyIle-----LysSerGlnGlnAsnMetTyrAla 336
Db 3994 CAGCGTGTACCAACCTATATTTGGGACAGTGTCAAGCAGAGCCAGCCACTCTGTGTTT 4053
QY 337 GlyAlaProAlaSerAlaAlaIleGlyThrAspIleGlyLysLeuProSerMetMet 356

```



```
Db 4054 GGACAGCAGCTAGTCAGTCTGCATTTGCTCTGGG----- 4089
QY 357 ProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArg 376
Db 4090 -----ACAGCTAATGCCAGTTCGGTTTCCCAATTTGGTAGCAGCACTACGAATTTC 4140
QY 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlnGly 396
Db 4141 AACTTCACAAACAATAATCATCAGCAGGTGTTCCATTTTGGTCCAGTCCCGCAGCGCT 4200
QY 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla 416
Db 4201 GCAGCTCAGCCAGCGCTCAGTTTCCAGGGTCTTTTCATTAGCCAGCTCTCCA---GCA 4257
QY 417 ArgPheThr 419
Db 4258 TCATTTTACA 4266

RESULT 35
CN762396/c 374 bp mRNA linear EST 20-MAY-2004
LOCUS ID0AAA4AB01RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA4AB01 5',
mRNA sequence.
ACCESSION CN762396.1 GI:47536319
VERSION EST.
KEYWORDS Acyrthosiphon pisum (pea aphid)
SOURCE Acyrthosiphon pisum
ORGANISM Acyrthosiphon pisum
REFERENCE 1 (bases 1 to 374)
AUTHORS Hunter, M., Martinez-Torres, D., Rahbe, Y., Sabater-Munoz, B.,
Stern, D., Tagu, D. and Winkler, P.
TITLE An expressed sequence tags database for the pea aphid Acyrthosiphon
pisum
JOURNAL Unpublished (2004)
COMMENT Contact: D. Tagu
INRA Rennes
UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France
Tel: +33.2.23.23.48.51.65
Fax: +33.2.23.23.48.51.50
Risk of contamination by bacterial sequences from obligatory
(Buchnera) or facultative endosymbionts. These sequences were
obtained in the frame of the International Consortium of Aphid
Genomics in collaboration with Genoscope
PCR Primers
FORWARD: CAGGAACAGCTATGACC
Plate: 4 row: B column: 1.
Location/Qualifiers
1. 374
/organism="Acyrthosiphon pisum"
/mol_type="mRNA"
/cultivar="developmentstage"
/db_xref="taxon:7029"
/clone="ID0AAA4AB01"
/tissue_type="whole insect"
/dev_stage="nymphs and adults (parthenogenetic females)"
/lab_hosts="XLI-Blue"
/clone_lib="ApMS"
/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI;
Sample name: ID0AAA; Plant growth place: Department of
Ecology & Evolutionary Biology, Princeton University;
Soil conditions: Soil; Sowing date: 01/06/1999;
Harvesting date: 01/06/1999; Stress date: no stress;
Description: Aphids inoculated on one-week old Vicia faba
under non-sterile conditions. All parthenogenetic stages
and both winged and wingless adults were collected for
library construction.; experimental condition: long
photopeiod (16-hr light/8-hr dark at 18 c)"

Db 4054 GGACAGCAGCTAGTCAGTCTGCATTTGCTCTGGG----- 4089
QY 357 ProIleGlnThrSerSerGlyAsnSerThrAlaArgAsnGlySerSerSerThrArgArg 376
Db 4090 -----ACAGCTAATGCCAGTTCGGTTTCCCAATTTGGTAGCAGCACTACGAATTTC 4140
QY 377 TyrSerGlnAspGlyTyrProGlnGlyAspLeuValAspValThrIleThrSerGlnGly 396
Db 4141 AACTTCACAAACAATAATCATCAGCAGGTGTTCCATTTTGGTCCAGTCCCGCAGCGCT 4200
QY 397 LysLeuGlnGlyLysTyrSerAsnSerGlnValValAspPheTyrAsnIleProLeuAla 416
Db 4201 GCAGCTCAGCCAGCGCTCAGTTTCCAGGGTCTTTTCATTAGCCAGCTCTCCA---GCA 4257
QY 417 ArgPheThr 419
Db 4258 TCATTTTACA 4266

RESULT 36
BG857482 348 bp mRNA linear EST 29-MAY-2001
LOCUS 1024052B06.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG857482
VERSION BG857482.1 GI:14238666
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE 1 (bases 1 to 348)
AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, B., Lefebvre, P.,
McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu
Location/Qualifiers
1. 348
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
Zap clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
```

ALIGNMENT SCORES:

ORIGIN

Alignment Scores: 0.0012 Length: 348
Pred. No.: 131.50 Matches: 34
Score: 46.67% Conservative: 15
Percent Similarity: 32.38% Mismatches: 43
Best Local Similarity: 5.09% Indels: 13
Query Match: 4 Gaps: 3
DB:

US-10-009-823A-1 (1-502) x BG957482 (1-348)

QY 10 ThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAlaAsnAla 29
Db 68 ACCGGATTAAACGCTGCGCCACCACTCGATGTTATGGCAACATATCGCAACTCC 127
QY 30 AenThrIleGlyTVrLysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeu 49
Db 128 GCCACTACGGCTTAAATCAGGCAGCGCTCTTTGCGATATGTTGCC----- 178
QY 50 AlalleGlySerThrGlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGly 69
Db 179 -----GGTTCC-----AAAGTGGGACTGGGGGTAAGAAAGTTGCC 211
QY 70 SerValArgThrIlePheThrGlnGlyAlaPheGluProGlyAsnSerValThrAspLeu 89
Db 212 GGTATCATCAGGACTTTACCGATGGCAGCACCAACCGCGGAGGTCTGGAGCTT 271
QY 90 AlalleGlyGlyLysGlyPhePheGln---ValThrLeuGluAspLysValHisTyrThr 108
Db 272 GGTATCAGCAGACGGTTTTTTCGTCGTGTAGACAGCACGCTTCGGTCTTCTACAGC 331
QY 109 ArgAlaGlyAsnPhe 113
Db 332 CGTAACGGACAAATTT 346

RESULT 37
AK089973
LOCUS 1723 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus skin RCB-1283 B16 melanoma cDNA, RIKEN full-length enriched library, clone:G430037K05 product:nucleoporin 98, full insert sequence.
ACCESSION AK089973
VERSION AK089973.1 GI:26105674
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE 11042159
PUBMED 10349636
REFERENCE 2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE 11042159
PUBMED 10349636
REFERENCE 3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
JOURNAL 5
REFERENCE 5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
JOURNAL 6
REFERENCE 6
ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., BONO, H., CARNINCI, P., FUKUDA, S., FURUNO, M., HANAGAKI, T., HARA, A., HASHIZUME, W., HAYASHIDA, K., HAYATSU, N., HIRAMOTO, K., HIRAKAWA, T., HIROZANE, T., HORI, F., IMOTANI, K., ISHII, Y., ITOH, M., KAGAWA, T., KASUKAWA, T., KATO, H., KAWAI, J., KOJIMA, Y., KONDO, S., KONNO, H., KODA, M., KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., MURATA, M., NAKAMURA, M., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OHSATO, N., OKAZAKI, Y., SAITO, R., SAITO, H., SAKAI, C., SAKAI, K., SAKAZUME, N., SANO, H., SASAKI, D., SHIBATA, K., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKIHARA, S., TAKEDA, Y., TANAKA, T., TOMARU, A., TOYA, T., YASUNISHI, A., MURAMATSU, M. and HAYASHIZAKI, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
Location/Qualifiers
1..1723
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6"
/db_xref="FANTOM,DB:G430037K05"
/db_xref="taxon:10090"
/clone="G430037K05"
/cell_line="RCB-1283 B16 melanoma"
/tissue_type="skin"
/tissue_lib="RIKEN full-length enriched mouse cDNA library"
/clone_lib="1500"
/misc_feature
/note="nucleoporin 98 (MGD|MG1:109404, GB|AF201386, evidence: BLASTN, 98%, match=580)
putative"
1703..1708
polya_signal
/note="putative"
1723
polya_site
/note="putative"
ORIGIN
Alignment Scores:
Pred. No.: 0.0148 Length: 1723
Score: 131.50 Matches: 122
Percent Similarity: 33.21% Conservative: 52
Best Local Similarity: 23.28% Mismatches: 205
Query Match: 5.09% Indels: 146
DB: 3 Gaps: 23
US-10-009-823A-1 (1-502) x AK089973 (1-1723)

QY 8 GlyAlaThrGlyMetLysThrHisSerThrGlyLeuGlyThrValSerAsnAsnIleAla 27
 DB 216 GGCAACATCTCAACATTGGGCAAAATCTGGCTTTGGTACACTAGTGGAGGACATT 275
 QY 28 AsnAlaAsnThrIleGly-----Tyr 34
 DB 276 GGAACATCTGCATTGGTTCTAGCAACAATACTGGAGGCTTATTGGAAATTCACAGACC 335
 QY 35 LysGlnGlnValValPheGlnAspLeuPheSerGlnAspLeuAlaIleGlySerThr 54
 DB 336 AAACAGGAGGATTATTTGGTACCATTTAGCCAGCCAGCAACCTCCACRAGCACT 395
 QY 55 GlySerGlnGlyProAsnGlnAlaGlyMetGlyAlaGlnValGlySerValArgThrIle 74
 DB 396 GGG-----TTTGGTTTGGCACATCAACAGGAATCAATCAATAGCTTA 437
 QY 75 Phe-----ThrGlnGlyAlaPhe----- 80
 DB 438 TTGTGAACCTGCAAGTACCGGACCACTCTTTCTCATCCAGAACATGCAATTTGCACNA 497
 QY 81 ---GluPro-----GlyAsnSerValThrAspLeuAlaIleGlyLysGlyPhe 96
 DB 498 AATAAACCAACTGGCTTTGGAAATTTGGNACCACTAGTACAGTGA-----GGACTC 551
 QY 97 PheGlnValThrLeuGluAspLysValHisTyrThrArgAlaGlyAsnPheArgPheThr 116
 DB 552 TTGTGAACCTACA-----AATACCACCTCTAATCTTTTGGTAGCACA 593
 QY 117 GlnAspGlyPheLeuAsnAspProSerGlyPheThrLeu-----MetGlySerArgIle 134
 DB 594 ---TCTGGCTCCCTTTTGGCCCAAGTAGTTTACAGCAGCACTACAGNACTACCATC 650
 QY 135 SerAsnAsnProAsnIleLysLysGluThrLeuGluProIleGlnLeuAspPheAsnAsp 154
 DB 651 AAATTTAATCT----- 662
 QY 155 ProThrValAlaLysSerProAlaLysThrSerThrAlaLeuAsnAlaValValAsnLeu 174
 DB 663 -----CCACTGGTACAGATCTATGCTCAAGCTGAGTGGAGTACACT 704
 QY 175 GlyAspSerThrAspLysThrGlnSerGluAlaAsnProTyrPheAlaLeuLeuGluSer 194
 DB 705 AACATCAGTACA-----AAGCATCAGTGTTACTGCTGTAAGAA 746
 QY 195 TrpLysGlyAsnGlyThrProProIleSerThrSerAsnTyrSerTyrAlaGlnProMet 214
 DB 747 TATGAAAGCAAGTCA-----TTAGAGGAACTA 773
 QY 215 ArgValTyrAspGlnGlnGlyAsnSerHisAspIleThrValTyrPheAspGlyAlaPro 234
 DB 774 CGTTTGGAGGATTATCAGGCTAACCG-----AAGGCCCAACAG 812
 QY 235 SerSerThrGlySerLysThrPheGluTyrLeuValAlaMetAsnProSerGluAspGly 254
 DB 813 AACCAAGTGGGAGGAGGCCACCGCTGCTTATTGGTCTTCTCCAGCA----- 863
 QY 255 SerAlaAlaSerGlyThrAspSerAlaGlyLeuLeuMetSerGlyThrMetThrPheSer 274
 DB 864 -----ACTTCCAGTGCACAGGGCTTTCAGCTCCTCCACCACTAATTC 908
 QY 275 SerAsnGlyGluLeuLysAsnMetThrAlaPheThrProThrGlySerAlaThrLys-As 294
 DB 909 GCCTTTTCATATGTCAGACAAACACTGCTTT-----GGAACCTAGCACAACTGGA 959
 QY 294 PheAsnAlaTrpGlnProAlaProLeu-----ValAsnGlyLeuProGlnPheSe 311
 DB 960 TTGTGAACAAATCCAGTGGTCTCTTTGGCCACAGAAATCAACAGACTACCACTCTCTTC 1019
 QY 311 rAlaAsnPheValGlyAlaGlyIleGln-----ProLeuThrLe 324
 DB 1020 AGCAAAACCAATTG-----GCAGGCTACACCCCGCAATATCGGCTTTCTTTGGTAAT 1076
 QY 324 uAspPheGlyIleLysSerGlnGlnAsn---MetTrpAla----- 336

DB 1077 ACCAGACCCCTTGGACAGCCAGCAACCACTATTTGGAGTAAACCAAGCC 1136
 QY 337 -----GlyAlaProAlaSerAlaAlaIleG1 346
 DB 1137 TCACCAACGAGGAGTCTTTTGGGACAGCTACAAACACAGCAGCTGGGACAGCATTTGG 1196
 QY 346 YThrAspIleGlyLysLeuProSerMetMetProIleGlnThrSerSerGlyAsnSerTh 366
 DB 1197 GACAGGAACAGGCTCTTTTGGGAGCCCAATACCTATTTGGTGGTGGTGGTGGTGGTGG 1256
 QY 366 rAlaAsnGlySerSerSerThrArgArgTyrSerGlnAspGlyTyrProGlnGlyAs 386
 DB 1257 GTTTGGCAATACAGCTTACAACTTTTGGAACCCAGCACACAGCTCTCTCTCATTTGG 1316
 QY 386 pLeuValaspValThrIleThrSerGluGlyLysLeuGlnGlyLysTyrSerAsnSerG1 406
 DB 1317 T-----ACACACAGTGGCGGCTCTTCCGGTTTGGCAACAAATACAG 1358
 QY 406 nValValaspPheTyrAsnIleProLeuAlaArgPheThrSerGluaspGlyLeuArgAr 426
 DB 1359 TGGAGCGAGTATTTTGGAGTAAGCCAGCAGCTTGGAACTTGGGAACCTGGACTT----- 1413
 QY 426 gGluGlyAsnAsnHisTyrSerAlaThrLeuAspSerGlyGlyProGluPheGlyLeuPr 446
 DB 1414 ---GGTACAGGATTTGGACAGCTGCCTATGAAATCAGAGGTTTTCATATTCTCTGGACC 1469
 QY 446 oGlyThrSerAsnTyrGlyLysLeuSerValAsnGlnLeuGluThrSerAsnValaspMe 466
 DB 1470 TGGAGTTACAGACTACTT-GCAGCCACCATCATGTGTGTGGAA---TCAGTGTGGATCT 1525
 QY 466 tSerArgGlu 469
 DB 1526 CCTGCAAGAG 1535

RESULT 38

AZ577764/c

LOCUS

DEFINITION

SP: NGR234 genomic clone 14e10, genomic survey sequence.

ACCESSION

AZ577764

VERSION

AZ577764.1

SOURCE

GSS.

KEYWORDS

Rhizobium sp. NGR234

ORGANISM

Rhizobium sp. NGR234

REFERENCE

Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.

1. (bases 1 to 324)

Genetic snapshots of the Rhizobium species NGR234 genome

Genome Biol. 1 (6), RESEARCH0014 (2000)

MEDLINE

21114532

PUBMED

11178268

COMMENT

Contact: Virginie Viprey

Laboratoire de Biologie Moleculaire des Plantes Superieures

University of Geneva

1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland

Tel: +44(0)1603450000

Fax: +44(0)1603450045

Email: virginie.viprey@bbsrc.ac.uk

Class: shogun.

Location/Qualifiers

1. .324

/organism="Rhizobium sp. NGR234"

/mol_type="genomic DNA"

/strain="ANU265"

/db_xref="taxon:394"

/clone="14e10"

/ANU265"

/notes="Vector: M13; derivative strain of NGR234 cured of pNGR234a"

ORIGIN

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 26, 2004, 09:04:16 ; Search time 40 Seconds
(without alignments)
1207.520 Million cell updates/sec

Title: US-10-009-823A-1
Perfect score: 2586
Sequence: 1 MMGSLFIGATGKTHSTGIG.....NSKSVTTADTMLOKALEKR 502
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	557.5	21.6	428	2 AG0219	flagellar hook pro
2	534.5	20.7	463	2 C71292	probable flagellar
3	527.5	20.4	462	2 F83510	flagellar hook pro
4	509	19.7	442	2 C70135	flagellar hook pro
5	498	19.3	403	2 AB0640	flagellar hook pro
6	498	19.3	403	2 S10365	flagellar hook pro
7	493	19.1	401	2 F90810	flagellar hook pro
8	493	19.1	401	2 B85670	flagellar biosynth
9	491.5	19.0	402	2 A64851	flagellar hook pro
10	485	18.8	718	2 G71888	flagellar hook pro
11	484.5	18.7	411	2 A11161	flagellar hook pro
12	484.5	18.7	411	2 A11520	flagellar hook pro
13	481	18.6	718	2 F64628	flagellar hook pro
14	435	16.8	482	2 E70460	flagellar hook pro
15	426.5	16.5	545	2 F81420	flagellar hook pro
16	418	16.2	413	2 AE0089	probable flagellar
17	413.5	16.0	405	2 E84969	flagellar hook pro
18	388	15.0	434	2 D82106	flagellar hook pro
19	384.5	14.9	865	2 E81271	flagellar hook cha
20	380.5	14.7	591	2 B87361	flagellar hook pro
21	370.5	14.3	263	2 A83956	flagellar hook pro
22	367.5	14.2	425	2 A12646	flagellar hook pro
23	367.5	14.2	425	2 H97428	flagellar hook pro
24	354.5	13.7	605	2 D64633	flagellar hook - H
25	349.5	13.5	605	2 G71881	probable flagellar
26	347	13.4	396	2 A83529	flagellar hook pro
27	333	12.9	264	2 G59622	flagellar hook pro
28	325.5	12.6	261	2 B72243	flagellar basal-bo
29	311	12.0	262	2 C70372	flagellar hook bas

30	307	11.9	261	2 H83510	flagellar basal-bo
31	303	11.7	260	2 H90810	flagellar basal bo
32	303	11.7	260	2 D85670	flagellar basal bo
33	303	11.7	260	2 C64851	flagellar basal bo
34	301	11.6	260	1 XEBBFG	flagellar basal bo
35	301	11.6	260	2 AD0640	flagellar basal-bo
36	291.5	11.3	423	2 E97165	flagellar hook pro
37	290.5	11.2	262	2 AC2644	flagellar basal-bo
38	290.5	11.2	262	2 C97426	flagellar basal-bo
39	288	11.1	262	2 F71801	flagellar basal-bo
40	287.5	11.1	454	2 S61455	flagellar hook pro
41	287	11.1	264	2 C71259	probable flagellar
42	286	11.1	262	2 A64718	flagellar basal bo
43	285	11.0	260	2 G84969	flagellar basal-bo
44	282.5	10.9	262	2 AE3645	flagellar basal-bo
45	274	10.6	261	2 AH0089	probable flagellar

ALIGNMENTS

RESULT 1

AG0219
flagellar hook protein FlgE [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004

C;Accession: AG0219

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; F
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, I

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AG0219

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-428 <KUN>

C;Cross-references: UNIPROT:Q82PB5; GB:AL590842; PIDN:CA090619.1; PID:gl5979825; GSPDB:GN

C;Genetics:

A;Gene: flgE

Query Match	21.6%	Score 557.5;	DB 2;	Length 428;
Best Local Similarity	29.1%	Pred. No. 2.1e-27;		
Matches 145;	Conservative 74;	Mismatches 196;	Indels 83;	Gaps 11;
QY	10	TKMTHSTGLGTVSNINIANANTIGYKQOQVVFQDLFQSDLAIGSTGSGGNQAGMGAQVG 69		
DB	8	SGMNAASNLVDVIGNNIANSATSGFKAGSVSFADMF-----GS-----QTGMGVKVA 55		
QY	70	SVRTTFTQGAPEPGNSVTDLAIGKGPFOVTLDEK---VHYTRAGNFRFTQDGFNDPSG 126		
DB	56	GTQDPNDGTATTNRRLDLAISONGFR--MQDSSGGIYARNGQFADENRNVNMQG 113		
QY	127	FTLMGSRISNNPNKKETLEPIQLDNDPTVAKSPAKTALNAVNLGSDTKTQSEAN 186		
DB	114	LNLGYPATGTPPTVQGANPVLSIQDMIS---AKATTSGNVANLTSTHVDIAEATS 170		
QY	187	PYFALLESKMGNGPPISTSNYSVAQPMRVYDQGNSHDITV-YFDCAPSTSGSKTEYL 245		
DB	171	PF-----DPNDPTYSFVNMTFTSLGNHEINVFYKRAEDATDGNWDVY 218		
QY	246	VAMPSEDSASGSDTSAGLMSGTMTFSNGELKNMTAFTPTGSAFKDLNAWQAPLVN 305		
DB	219	TRDSSAKVTDPAEDTDPAAAAKRGSMVFDNGALKNVTNGT-NATSTDTFTTIPMGVNV 277		
QY	306	GLP--QFSANFVAGIQPLTLDGKISQOQNMWAGAPASAAAIGTDIGKLSMMPIQISSG 363		
DB	278	GAPQSFALNVAGS-----KQKN----- 295		
QY	364	NSTANGSSSTRYSQDGYPGQDLDVDTITSEGLKQKYSNSQVDFYNTPLARFTSEDG 423		
DB	296	-----TGADSVIAQNTGYAAGEFTGQINSQDGVGTVGTSNQQTQLLGQIVMVFNSPEG 350		

QY 424 LRREGNHYATLDSGPEFLPGTSNYKLSVNOLETNSVDMSEMNMIHQGFQMN 483
Db 351 LSEGNVWVKETQSSGNFTLTAGSGGFTLTSGALESSNVDSKELVNMIVAQRYQSN 410
QY 484 SKSVTTADTLMQALELK 501
Db 411 AQTIKTDQILQTLVSLR 428
RESULT 2
C71292
probable flagellar hook protein (flgE) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004
C:Accession: C71292
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: C71292
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-463 <COL>
A:Cross-references: UNIPROT:Q07884; GB:AE001244; GB:AE000520; NID:G3323005; PIDN:AA66569
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0727

Query Match 20.7%; Score 534.5; DB 2; Length 463;
Best Local Similarity 28.1%; Pred. No. 6.2e-26;
Matches 154; Conservative 80; Mismatches 183; Indels 131; Gaps 14;
QY 1 MMSLFGATGKTHSTGLTVSNINANANTIGYKQVQVFDLFSQDLAIGS-----TG 55
Db 1 MMSLFGVSGNQHQHTRDVIGNVANVTGFRGVNFQDLISQOLSAARPNEEVG 60
QY 56 SQGNQAGMGAVQSVRTITQGAPEFGNSVTDLAIGKGFQVLTEDKVHYTRAGNFRP 115
Db 61 GVNPEVGLVLIASIDVHTQALQTTGINTDVSIOGSGFFVLKSGEKTFFRAGAFV 120
QY 116 TQDGLNDPS-----GFTLMGSRISNNPNTKKTLEPIQLDFNDPTVAKSPAKTST 166
Db 121 DNAGTLVNPANGVRQGMADVAGERLINSQAQDLVIPG-----QKIDAQOTS 172
QY 167 ALNAVAVNLGSDTKTQSEANPYFALLESWKNGTPPSTNSYVAQPMRVYDQGNHSDI 226
Db 173 TVHYACNLKDELPELAADANEADVRSKSTW-----TTDF-----QYVDSFGQOHTL 217
QY 227 TVYFDGAPSTGKTFYLVAMP-----SDGSA-----ASGT 260
Db 218 QINFSRVPGT--NNQWQATVADVDEVDQTRVGVGTSDGAANTFVNFDFGHASVT 275
QY 261 DSAGLLMSGTTFSSNGELKNMTAFTPTGSATKDLNAWQAPLVNGLPQFSANFVCGAGIQ 320
Db 276 DTAG-----NVTG--PTGQV-----LLEASYDVUGA-----299
QY 321 PLTLDFGIKSOQNMWAGAPASAAIGTDIGKLPMPPIQTSNGSNSTARN-----GSSST 374
Db 300 -----NPDAGQVTRHAFTLNLGEI-----GTARNITITQFAERSTT 335
QY 375 RRYSDQGVPOGLDVTITSEGLQKGYNSQVDFNIPLARFTSEDLRREGNNHYSA 434
Db 336 KATQDQYANGYLENFKIDQSGVITGVYNGVSDIGQALAGAFANQGGLEKAGENTYVQ 395
QY 435 TLDSGGPEFLPGTSNYKLSVNOLETNSVDMSEMNMIHQGFQMNKSVTTADTLM 494
Db 396 SNNSGIANISTGVMGKLIAGTLEMSNVLDLQFTDMITQKGFQAGAKTITSDTLM 455
QY 495 QKALELKR 502
Db 456 DTVLSLKR 463

RESULT 3
F83510
flagellar hook protein flgE PA1080 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: F83510
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathos
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83510
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-462 <STO>
A:Cross-references: UNIPROT:O914P9; GB:AE004539; GB:AE004091; NID:G9946985; PIDN:AAG04465
A:Experimental source: strain PA01
C:Genetics:
A:Gene: flgE; PA1080
Query Match 20.4%; Score 527.5; DB 2; Length 462;
Best Local Similarity 28.5%; Pred. No. 1.7e-25;
Matches 150; Conservative 84; Mismatches 199; Indels 93; Gaps 16;
QY 4 SLFIGATGKTHSTGLTVSNINANANTIGYKQVQVFDLFSQDLAIGTSGSQGNQAG 63
Db 2 SFIGLGIQAASSGLNVTGNINAGTVGFKQSEAFADYVAAV-LGS-----GSNPGQ 56
QY 64 MGAQVGSVRTITQGAPEFGNSVTDLAIGKGFQVLTEDKVHYTRAGNFRPDTODGFLND 123
Db 57 SGLVLSDSVQSFQKQGNIDSTNSVLDAINGGFFVTSNNGAISYTRAGYFNTDQDFIVD 116
QY 124 PGFTLMGSRISNNPNTKKTLEPIQLD-FNDPTVAKSPAKTSTALNAVNLGSDTKTQ 182
Db 117 NNGYRLOQYAVGPNQQLGVVVDLKVERNQAPQATSSIQSYNLNLTSLK-----P 168
QY 183 SEANPYFALLESWKNGTPPSTNSYVAQPMRVYDQGNHSDITVYF---DGAPSTP-- 237
Db 169 PVTVPF-----DPSDAATYSSSLGIYDSQNSHTWSQFFIKNEPDPNATPP 216
QY 238 ---GSKTFEYLV-AMNPSEDSGAAGTDSAGLLMSGTTFSSNGELKNMTAFTPTGS--- 290
Db 217 IPENSWTMKVLIDGVNPLDPSNKT-----NSFNVTFDASQMTSVRA--PDGSTSG 266
QY 291 -----ATKDLNAWQAPLVNGLPQFSANFVCGAGIPLTLDFGIKSOQNMWAGAPASAA 344
Db 267 PGFSIDATNVIQSPA---TGNPPTP---GTGWIPAASD-GKTPPTYAWNGATGAASG 318
QY 345 IGTDIGKLPMPPIQTSNGSNSTARNSSSTRYS-----QDGYPOGDLVDVITSE 395
Db 319 ISFDMRK-----TTQYSTAFAPQSNPIODGYTTQGLAGLEIDDT 356
QY 396 GKLOKGYNSQVDFNIPLARFTSEDLRREGNNHYSATLDSGPEFLPGTSNYKLS 455
Db 357 GVIFARYTNGSKVQGVVLANFANIQLTPIGKTSVWQSESEGEPAVGAPRSCTLGALQ 416
QY 456 VNQLETSNVDMSEMNMIHQGFQMNKSVTTADTLMQALELK 501
Db 417 SGALASNVDISNELVNLVHQRNYQNAKTIQTEDAVTQTINLR 462

RESULT 4
C70135
flagellar hook protein (flgE) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: C70135
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: C70135
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-442 <KLE>
A;Cross-references: UNIPROT:Q44767; GB:AE001137; GB:AE000783; NID:92688160; PIDN:AAC6666
A;Experimental source: strain B31

Query Match 19.7%; Score 509; DB 2; Length 442;
Best Local Similarity 28.2%; Pred. No. 2.3e-24;
Matches 150; Conservative 71; Mismatches 190; Indels 120; Gaps 13;
QY 1 MMGSLFATGKTHSTGLTGVNNIANANTICYKQOQVVFQDLFSDLAIGS-----TG 55
DB 1 MMSLYSGVGLQHQTDMVDVGNVNIANVNTICFKKGRVNFQDMISQISGASRPDARG 60
QY 56 SQGNQMGAGVGVRTITQAFPGNSVTDLAIGKGFQFVTLKDVHYTRAGNFRF 115
DB 61 GTNPKQVGLGMNVASDITHTQAFQSTQKASDLGVSGNGFFILKEGKNLFYTRAGFDV 120
QY 116 TDGFLNDPQGTLMGSRISN--NPNIKKTLEPIQLDFND---PTVAKSPAKTSTALNA 170
DB 121 DSRHLNVAN---GNRIQHWARDLEGEKVINTASDIEDLIPIGDKGAKSTKNVTF 176
QY 171 VVNLGSDTKQSEANPYFALLSWKNGTTPPISTSNYSVAQPMRVYDQGNSHDITVYF 230
DB 177 ACNLDKRLPIQEGANPADTARGTWVVKNS-----LYDSFGN---VSVL- 217
QY 231 DGAPSSGTSTFEVYLVAMVSEDSASGSDSAGLLMSGTMTFSSNGELKNTAFTPTGS 290
DB 218 -----ELR----- 220
QY 291 ATKQLNA---WQAPLVNGLPQFSANFVAGIOLPLTLP-----GIKQQQNWAG 337
DB 221 VKDLNTPNLWNTVLING--EQNSF-----TLGFDNEGALASLNGCPQKGDILQ 270
QY 338 APASAAAIGTDIGKLPSPMPIQTSSG-----NSTARNSSSTRYSQDGYQGLVDV 390
DB 271 IPITFNVLGANVGEVGEQTVNLKLTGVSYTSITQFADSSSKAIIDQGYGMWENY 330
QY 391 TITSEGLQKYSNVQVDFYNIPLARFTSEGLRREGNNHYSATLDSGGPFGLPQTSN 450
DB 331 EIDQGVIVGYNGIRDLGKIALASFPNPGGLAKSGDINFVETSSNGQVRIGETGLAG 390
QY 451 YGKLSVNOLETNSVDMGEMVMIIIOGFQOMNSKSVTTADTMLOKALELK 501
DB 391 LGDIRSGVLEMANVDLAEQFTDMIVTQGFQANAKTITTSQDLQLQELVRLK 441

RESULT 5
AB0640
flagellar hook protein FlgE [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhimurium
C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A;Note: This species has also been called *Salmonella typhi*
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0640
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhimurium
A;Reference number: AB0640; MUID:21534947; PMID:11677608
A;Accession: AB0640
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-403 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08301.1; PID:g16502347; GSPDB:GN00176
C;Genetics:
A;Gene: STY1216

Query Match 19.3%; Score 498; DB 2; Length 403;
Best Local Similarity 28.7%; Pred. No. 9.5e-24;
Matches 144; Conservative 76; Mismatches 167; Indels 114; Gaps 14;
QY 10 TGMKTHSTGLTGVNNIANANTICYKQOQVVFQDLFSDLAIGS-----TG 69
DB 8 SGLNAATNLVDVGNVNIANVNTICFKKGRVNFQDMISQISGASRPDARG 55
QY 70 SVRTITQAFPGNSVTDLAIGKGFQFVTLKDVHYTRAGNFRFQDGLNDPQGT 128
DB 56 GITQDFTDGTITNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKLDENLNMQMQ 115
QY 129 LMGRISNNPIKKTLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTKQSEANPY 188
DB 116 LTGYPATCTPTTIQGANPAPITIPNTLMA---AKSITTASMQNLN-NSTDPVPSKT--- 168
QY 189 FALLSWKNGTTPPISTSNYSVAQPMRVYDQGNSHDITVYDGPAPSTGSKTPEYL 245
DB 169 -----PFSVSDADSYNKKGTVTYDSCGNADHNMVYF-----VKTQDNEWA 209
QY 246 VAMNPSEDSASGSDSAGLLMSGTMTFSSNGELK-----NMTAFTPTGSKATKOLNAWOP 300
DB 210 VTTHDSSDPATAPTTA-----STLKENENGILSGGTVNTITGTINGATA----- 256
QY 301 APLVNLGDPQFSANFVAGIOLPLTLDGFIKSKQONMWAGAPASAAAIGTDIGKLPSPMPIQT 360
DB 257 -----ATFSLPLNS-----MQON-----T 271
QY 361 SSGNSTARNGSSSTRYSQDGYQGLVDVITSEGLQKYSNVQVDFYNIPLARFTS 420
DB 272 GANNIVATN-----QNGYKPGDLVSYQIINNDDGTVVGNYSNEQVQLQIVLVANFAN 322
QY 421 EDGLRREGNNHYSATLDSGGPFGLPQTSNYSKLSVNOLETNSVDMGEMVMIIIOGF 480
DB 323 NEGLASQGDVNWAAQASVALLGTAGSGNFKLTNGALEASNVDSLSELVNMIVAQSNY 382
QY 481 QNNSKSVTTADTMLOKALELK 501
DB 383 QSNAQTIKTQDQILNLTNLNLR 403

RESULT 6
S10365
flagellar hook protein flgE - *Salmonella typhimurium*
C;Species: *Salmonella typhimurium*
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C;Accession: S10365; S15357
R;Homma, M.; DeRosier, D.J.; Macnab, R.M.
J. Mol. Biol. 213, 819-832, 1990
A;Title: Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their
A;Reference number: S10361; MUID:90294298; PMID:2193164
A;Accession: S10365
A;Molecule type: DNA
A;Residues: 1-403 <HOM>
A;Cross-references: UNIPROT:P16322; EMBL:X51737; NID:G47671; PIDN:CAA36022.1; PID:G47673
R;Jones, C.J.; Macnab, R.M.; Okino, H.; Alzawa, S.I.
J. Mol. Biol. 212, 377-387, 1990
A;Title: Stoichiometric analysis of the flagellar hook-(basal-body) complex of *Salmonella typhimurium*
A;Reference number: S15353; MUID:90204563; PMID:2181149
A;Accession: S15357
A;Molecule type: protein
A;Residues: 2-5 <JON>
C;Genetics:
A;Gene: flgE
A;Map position: 23 min
C;Keywords: basal body; flagellum
F;2-403/Product: flagellar hook protein #status experimental <MAT>

Query Match 19.3%; Score 498; DB 2; Length 403;
Best Local Similarity 28.7%; Pred. No. 9.5e-24;
Matches 144; Conservative 76; Mismatches 167; Indels 114; Gaps 14;

368 RNSSTRRYSQDGYPGDLVDVITSEGLKQKYNSQVDFYNIPLARFTSEDLRRE 427
277 T-----TQNGYKPGDLVSYQINDDGTGVNGYNEQTQLLGGQIVLANFANNEGLASE 327
428 GNNHYSATLSDGCGPECLPCTSNYKLSYNQLETSNVDMRSVMNMIILQRFQMNKSV 487
328 GDNVWSATQSSGVALIGTAGTNGFGLTNGALERSVNDLSKELVNNIVQRYQSNQRII 387
488 TTADTLMQKALELK 501
388 KTQDQILTLVNL 401
RESULT 9
A:Accession: A64851
A:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: A64851
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64851
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-402 <BLAT>
A:Cross-references: UNIPROT:P75937; GB:AE000208; GB:U00096; NID:G1787308; PIDN:AACT4160.
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: flgE; flak; fla FV
C:Keywords: basal body; flagellum
F:2-402/Product: flagellar hook protein flgE #status predicted <MAP>
Query Match 19.8%; Score 491.5; DB 2; Length 402;
Best Local Similarity 28.8%; Pred. No. 2.4e-23;
Matches 143; Conservative 75; Mismatches 173; Indels 105; Gaps 15;
10 TGMKTHSTGLTVSNNIANTIGYKQVQVFDLFSQDLAIGTSQGNQAGMAQV 69
8 SGLNAAATNLVDIGNNIANSATYGFKSGTASFAFMFA-----GS-----KVGLGVKVA 55
70 SVRTIFQGAPEPGNSVTDLAIGKGFQ-VTLEDKHYTRAGNFRFTQDGLNDPSGFT 128
56 GITQDFTDGTNTGRGLDVAISQNGFFRLVDSNGSVFYSRNGQFKLDENRNLVNMQLQ 115
129 LMGSRISNNPIKETEPIQLDNDPDTVAKSPAKTSALNAVNLGSDTDKTOSEANPY 188
116 LTGYPATGTPPTIOGGANPTNISIPNTLMA---AKTTTASMQINL-NSSDPLPT-----166
189 FALLESMKNGTPTPISTSN---YSVAQPMRVYDQGNSHDITVYFDGAPSGSTGKTPYL 245
167 -----VTFFSASNADSYNKKSVTVFDGQNAHDMSVF-----VKTGDNWQ-- 209
246 VAMPPESDGSAAGTDSAGLMSGTMTFFSNGELKNNMTAFPTTGSATKDLNAMQAPLVN 305
210 VVTQSDSPNSIAKT-----ATTLEFNANGTL--VDGAMANNIATGAINGAEP-----256
306 GLPQSANFVAGIOPLTLDGFKISQONWAGAPASAAAIGTDIGKLPSPMPIOTSSGNS 365
257 ---TFLSFLNS-----MQQN-----TGANNI 275
366 TARNGSSSTRYSQDGYPGDLVDVITSEGLKQKYNSQVDFYNIPLARFTSEDLR 425
276 VAT-----TQNGYKPGDLVSYQINDDGTGVNGYNEQTQLLGGQIVLANFANNEGLA 326
426 REGNNHYATLSDGCGPEGLPCTSNYKLSYNQLETSNVDMRSVMNMIILQRFQMNKSV 485
327 SEGDNVWSATQSSGVALIGTAGTNGFGLTNGALERSVNDLSKELVNNIVQRYQSNQ 386
486 SVTTADTLMQKALELK 501

387 TIKTQDQILTLVNL 402
RESULT 10
G1888
flagellar hook protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: G1888
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-718 <ARN>
A:Cross-references: UNIPROT:Q92KY0; GB:AE001509; GB:AE001439; NID:G4155350; PIDN:AAD06361
A:Experimental source: strain J99
C:Genetics:
A:Gene: flgE
Query Match 18.8%; Score 485; DB 2; Length 718;
Best Local Similarity 25.0%; Pred. No. 1.4e-22;
Matches 181; Conservative 94; Mismatches 222; Indels 226; Gaps 22;
1 WMGSFIGATGKTHSTGLTVSNNIANTIGYKQVQVFDLFSQDLAIGS-----TGS 56
1 MURSLWSGVNGMAHQAIALDIENNIANNVNTGFKYASAFVDMLSQVKLIATATAPYKNGL 60
57 QGPN--QAGMAQVGSVRTIFQGAPEPGNSVTDLAIGKGFQVFDLTKV--HYTRAGN 112
61 AGQNDPSVGLGVGDATTIKFSQGNQNTDKTDLAIQDGFIIISPDRTGRNTRDGE 120
113 RFTQDGLNDPSGFTLM-----GSRISNNPIKETEPIQLDNDPDTVAKSPAK 163
121 FLFDQSGSLVTTGGVLVQGVWVNGSDTKGKSDTDLKVDNTGPLENTRIDPMV-MEAR 179
164 TGT--ALNAVNLGSDTDKTOSEANPYFALLESWK--GNGTPISTSNYSYAQPMR----215
180 ASNRISMRANLNRHADQTA-----VFALDSSAKTPSDGINFVYDGTNLQAQVAEDMGS 235
216 VYDQGN-----SHDITVYFDGA-----PS-----STGSKTFEY 244
236 LYNEGDGALLNENQGIWVSYKSAKVXKIDLPSAENSTLELNGVKISFTNDSAVRTSSL 295
245 LVAMPPESDGSAAG-----TDSAGLLMSGTMTFFSNGELKN-----281
296 VAAKNAINAVKSGTGLEAVLDGKQLENTNLDGDEKLNIVVTQAGTGAFFANFLDGDK 355
282 -MTAF-----TPTGSATKDLNAMQAPLVNGLPQFSANFV-----GAGIQPLTL 324
356 DVTAFKYSYTHSISPNADIGQFRTTDLRALIHDANIVKDPISLADNYQDSASIGVTIN 415
325 DFG-----TKSQNNW-----AGAPASAA 343
416 QYGMFEINKNKNNVTKENLINFVSGYSDSVTNVNVFKVAMKGLNTASLIEGASASS 475
344 -----AIGT-----347
476 KPTHATHATSIDVIDSLGTHAMRIEYRSGGAEMNFRVIVPEPGLVGGSAARPNVFE 535
348 -----DIGKPSMMP-----IQSSGNSTARNSSSTRYS-----Q 379
536 GRUHFVNDGSLAGMNPFLQFPKNGADAPQRINIAFGSSGFDGLTSDVKLSETVAIQ 595
380 DGYPOQDLVDVITSEGLKQKYNSQVDFYNIPLARFTSEDLRREGNNHYATLSDG 439
596 NGQAGDLMDVRDSDGVLLGAFSNGRTLLAQVALANFANDAGLQALCGNVFSQTGNSG 655
440 GPBFGPLPGTSNYKLSYNQLETSNVDMRSVMNMIILQRFQMNKSVTTADTLMQKALE 499

Db 656 QALIGAANTGRGSGKLESSNVDLSRLNLIVVQGFQANSKAVTTSDOILNTLLN 715
QY 500 LXR 502
Db 716 LKQ 718
RESULT 11
AII161
flagellar hook protein FlgE homolog lmo0697 [imported] - Listeria monocytogenes (strain
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AII161
R/Claser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
Science 294, 849-852, 2001
A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative Genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AII161
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-411 <GLA>
A/Cross-references: UNIPROT:Q92DV7; GB:NC_003210; PIDN:CAC98775.1; PID:g16410086; GSPDB:
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo0697

Query Match 18.7%; Score 484.5; DB 2; Length 411;
Best Local Similarity 27.0%; Pred. No. 6.8e-23;
Matches 137; Conservative 77; Mismatches 188; Indels 105; Gaps 11;
QY 1 MMGSLFIGATGKTHSTGLGTVSNINANTIGYKQOVVFDLFSQD---LAIGSTGS 56
Db 1 MNQMTYTAISGNARFQQAALSVTSNNINANTIGYKQSVFNDLLYQNTMGSVAGGLYAG 60
QY 57 QGPNOAGMAGVGSVRTITTCQAFEPGNSVTDLAIGGKGFQV--TLEDKVHYTRAGNFR 114
Db 61 TNPMSFGSGKIGAILTDYTAGSPTSGRNKDAALQGRFFIAGDNAGNIVVTRDGSFA 120
QY 115 FTQDGLNDPSGFTLMGSRISNNENIKETLEPIOLDPNDPTVAKSPAKTSTALNAVNL 174
Db 121 VSDNNLTQOQKYVMGYATDKNGVNLGNLQPIPLNSAIPGEATKNGSLSGNIPLDW 180
QY 175 GDSTDKTQSEANPYFALLESWKNGTTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGA- 233
Db 181 GEK-DTISSE-----LSVYDNAGGKHLQVNMKAAT 210
QY 234 PSSSTGSKTFEYLAVNPSDEGSAASCTDSAGLLMSGTMTFSSNGELKMTAFTPTGSATK 293
Db 211 PDASGNVSYEIQM---DGKALTPP-----VTGLTNAOQELTNPDA----- 251
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTLDLFGIKSQQNNWAGAPASAAAIGTDIGKLP 353
Db 252 -LKNIQINSTVNG-----KQVNM-----GLNLSGLT 276
QY 354 SMPQTSSGNSTARNSSSTRYSQDGYPOGDLVDVTITSEKLGKQKYSNSQVDFYNI 413
Db 277 NY-----GTNQVFSPTSGKGAATVKDYAVTDSGYIAVSYSDDGVIPIVAOL 322
QY 414 PLARFSDGLRRGNHYSATLDSGGPEGLPQTSNYGKLSVNOLETNSVNDMSREVMNM 473
Db 323 AVATFSNEDGLVKMGNGEYVPGSSGDVYGVAGQNGAGGAGGSSLEGSNVDLSREFVNL 382
QY 474 IIRQGFQMSKSVTTADTLMQKALEL 500
Db 383 MTYQSGFQGNKTVIRVADDDVMKQIVNL 409
RESULT 12

AII520
flagellar hook protein FlgE homolog lin0705 [imported] - Listeria innocua (strain Clip11;
C/Species: Listeria innocua
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AII520
R/Claser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
Science 294, 849-852, 2001
A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative Genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AII520
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-411 <GLA>
A/Cross-references: UNIPROT:Q92DV7; GB:AL592022; PIDN:CAC95937.1; PID:g16413157; GSPDB:GN
A/Experimental source: strain Clip11262
C/Genetics:
A/Gene: lin0705

Query Match 18.7%; Score 484.5; DB 2; Length 411;
Best Local Similarity 27.0%; Pred. No. 6.8e-23;
Matches 137; Conservative 77; Mismatches 188; Indels 105; Gaps 11;
QY 1 MMGSLFIGATGKTHSTGLGTVSNINANTIGYKQOVVFDLFSQD---LAIGSTGS 56
Db 1 MNQMTYTAISGNARFQQAALSVTSNNINANTIGYKQSVFNDLLYQNTMGSVAGGLYAG 60
QY 57 QGPNOAGMAGVGSVRTITTCQAFEPGNSVTDLAIGGKGFQV--TLEDKVHYTRAGNFR 114
Db 61 TNPMSFGSGKIGAILTDYTAGSPTSGRNKDAALQGRFFIAGDNAGNIVVTRDGSFA 120
QY 115 FTQDGLNDPSGFTLMGSRISNNENIKETLEPIOLDPNDPTVAKSPAKTSTALNAVNL 174
Db 121 VSDNNLTQOQKYVMGYATDKNGVNLGNLQPIPLNSAIPGEATKNGSLSGNIPLDW 180
QY 175 GDSTDKTQSEANPYFALLESWKNGTTPPISTSNYSYAQPMRVYDQGNSHDITVYFDGA- 233
Db 181 GEK-DTISSE-----LSVYDNAGGKHLQVNMKAAT 210
QY 234 PSSSTGSKTFEYLAVNPSDEGSAASCTDSAGLLMSGTMTFSSNGELKMTAFTPTGSATK 293
Db 211 PDASGNVSYEIQM---DGKALTPP-----VTGLTNAOQELTNPDA----- 251
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTLDLFGIKSQQNNWAGAPASAAAIGTDIGKLP 353
Db 252 -LKNIQINSTVNG-----KQVNM-----GLNLSGLT 276
QY 354 SMPQTSSGNSTARNSSSTRYSQDGYPOGDLVDVTITSEKLGKQKYSNSQVDFYNI 413
Db 277 NY-----GTNQVFSPTSGKGAATVKDYAVTDSGYIAVSYSDDGVIPIVAOL 322
QY 414 PLARFSDGLRRGNHYSATLDSGGPEGLPQTSNYGKLSVNOLETNSVNDMSREVMNM 473
Db 323 AVATFSNEDGLVKMGNGEYVPGSSGDVYGVAGQNGAGGAGGSSLEGSNVDLSREFVNL 382
QY 474 IIRQGFQMSKSVTTADTLMQKALEL 500
Db 383 MTYQSGFQGNKTVIRVADDDVMKQIVNL 409

RESULT 13
F64628
flagellar hook protein flgE - Helicobacter pylori (strains CCUG 17874, NCT11637, 26695)
C/Species: Helicobacter pylori
C/Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C/Accession: F64628; S78522; S61442; S61456
R/Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney
son, J.D.; Kelley, J.K.; Cotton, M.D.; Weidman, J.N.; Fujii, C.; Bowman, C.; Wattey, L.
Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64628
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-718 <COM>
A;Cross-references: UNIPROT:P50610; GB:AB000596; GB:AB000511; NID:G2313982; PIDN:AA00791
R;O'Toole, P.W.; Kostrzynska, M.; Trust, T.J.
submitted to the EMBL Data Library, May 1994
A;Reference number: S78522
A;Accession: S78522
A;Molecule type: DNA
A;Residues: 1-105; 'N', 107-236, 'Y', 238-258, 'P', 260-372, 'NS', 375-412, 'TI', 415-508, 'D', 510-
A;Cross-references: EMBL:U09549; NID:G600723; PIDN:AAA92803.1; PID:G600725
A;Experimental source: CCUG 17874; NCTC 11637
R;O'Toole, P.W.; Kostrzynska, M.; Trust, T.J.
Mol. Microbiol. 14, 691-703, 1994
A;Title: Non-motile mutants of *Helicobacter pylori* and *Helicobacter mustelae* defective in
A;Reference number: S61442; MUID:95198543; PMID:7891557
A;Accession: S61442
A;Molecule type: DNA
A;Residues: 1-83; 85-105, 'N', 107-236, 'Y', 238-258, 'P', 260-372, 'NS', 375-412, 'TI', 415-508, 'D'
A;Cross-references: EMBL:U09549
A;Experimental source: CCUG 17874; NCTC 11637
A;Note: only a part of the nucleic acid sequence is shown
A;Accession: S61456
A;Molecule type: protein
A;Residues: 1-14, 'X', 16-25 <OZO>
A;Experimental source: CCUG 915
C;Genetics:
A;Gene: flgB
C;Keywords: flagellum

Query Match 18.6%; Score 481; DB 2; Length 718;
Best Local Similarity 24.6%; Pred. No. 2.5e-22;
Matches 178; Conservative 96; Mismatches 223; Indels 226; Gaps 21;

Qy 1 MMSGLFATGKTHSTGLTVSNINANTIGYKQVQVFDLFSQDLAIGS-----TGS 56
Db 1 MLRLWGVNGVGMQAHQALDIESNINAVNTTGFKYGRASFDVMSQVKLIATAPYKNGL 60
Qy 57 QGPN--QAGMAQVGSVRTIFTQGAFFPGNSVTDLAIGKGFFQVTDLKV--HYTRAGN 112
Db 61 AGQDNFVGLGVGDVATTKIFSQNGIQTNDVKTDLAQGDGFFIISPDGRITRNFEDGE 120
Qy 113 FRFTQDGLNDPDSGFTLM-----GSRISNNPNIKETLEPIQLDFNDPTVAKSPAK 163
Db 121 FLFDQSGLVTTGGLVQGWVRNGSDTGNKGSDDALKVDNTGPLENIRIDPGMV-MPAR 179
Qy 164 TST--ALNAVNLGSDTKTQSEANPYFALLESNK--GNGTPISTSNYSYAO----- 212
Db 180 ASNRISMPANLNGRHADQTA--IPALDSSAKTSPDGINPVDYSGTNLAQVAEDMGS 235
Qy 213 -----PMRYVDDQ-----NSHDITVYPDGAP-----SSTGSKTFEY 244
Db 236 LCNEDGALLNENGIWVYSKAKWVDILPSAENSTLELNGVKISFTNDSAVSRSSSL 295
Qy 245 LVAMPNSEDGGAAG-----TTSAGLLMGTTFFSNGELKN----- 281
Db 296 VAAKNAINAVKSQTGIEAYLQKQLRLNTNELDGDEKLNIVVQTAGTAPANFLDGDK 355
Qy 282 -WTAF-----TPTGSAKDNLNAQWAPLVNGLPQFSANFV-----GAGIQPLTL 324
Db 356 DVTAFKYSYTHSISPNADIGQFRITEDLRALIQHDANIVKDPDLADNVQDSAAISGVSN 415
Qy 325 DFG-----IKSQQNMW-----ACAPASAA 343
Db 416 QYGMPEINKNKKNVKENLIFVSGYSSDVTNNVLFNKNAMKGLNTASLIEGGASASS 475
Qy 344 -----AIGT----- 347
Db 476 KETHATHATSDVIDSLGKHAIRIEFYRSGGAENFRVIVPEPGLVGSAAARNVPEG 535

Qy 348 -----DIGKLPMMMP-----IOTSSGNSTARNSSSTRYS-----Q 379
Db 536 GRLEHFNNDGSLAGNPPPLLPDPKNGADAPORINLAFGSSGFDGLTSDVTKISETAYAEQ 595
Qy 380 DCFPGDLVDVTIIESEKLGKQKYSNSQVDFYNIPLARFTSEDLGLRRGNHYSATLDSG 439
Db 596 NGYQAGDLMDFRFDGVDGVLGAFSNGRGTALAQAVALANFANDAGLQALGCVFSQTSNGS 655
Qy 440 GPEFGLPGTNSYKLSVNVQLETSNVDMREVMNMIIOQGFQMSKSVTTADTLMLOKALE 499
Db 656 QALIGAANTGRRGSISSGKLESSNVDLRSRLNLIVQGFCAKSAVTTSDQILNTLIN 715
Qy 500 LKR 502
Db 716 LKQ 718

RESULT 14
E70460
flagellar hook protein FlgE - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: E70460
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
A;Reference number: A70300; MUID:98196666; PMID:9537320
C;Accession: E70460
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-482 <AQF>
A;Cross-references: UNIPROT:O67711; GB:AB000760; NID:G2984138; PIDN:AAC07675.1; PID:G2984
A;Experimental source: strain VFS
C;Genetics:
A;Gene: flgE

Query Match 16.8%; Score 435; DB 2; Length 482;
Best Local Similarity 27.6%; Pred. No. 1e-19;
Matches 143; Conservative 94; Mismatches 204; Indels 78; Gaps 23;

Qy 1 MMSGLFATGKTHSTGLTVSNINANTIGYKQVQVFDLFSQDLAIGS-----STGSQ 57
Db 1 MLRSFYNAITGMDVSRFALDVTSDNLANANTVGFKKSRPIFQDMVSO-VVVGILNTTGTV 59
Qy 58 GPNQAGMAQVGSVRTIFTQGAFFPGNSVTDLAIGKGFF--QVTDLKVHYTRAGNPRF 115
Db 60 KTTTFAGAVVDSVTKQVTFIGSKQTEITTDLAIGKALFILRDVLTNGYTYTRDGRFPI 119
Qy 116 TDGFLNDPSGFTLMGSRISNNP--NIKKTLEPIQLDFNDPTVAKSPAKTSTALNAV 172
Db 120 NREGYLINGLYVQGVFKV--NPVTGEVGTGTQLEDIRVETQIPPKATG----EIVFPNPT 173
Qy 173 NLGSDTKTQSEANPYFALLESKNGTTPISTSNYSYAOQPMRVYDQGNSHDITVYFDG 232
Db 174 NLDERAPIIDQTTTP-----NPLDSFTYNYRYTLTIYDLSLGREVPADIYF-- 219
Qy 233 APSSTGSKTFE-YLVA-----MNPSEDGSAASGTDAGL-LMSGTMTFSSNGELKNMTPA 284
Db 220 --VKGTGNQWVYFLASLXERYINDWNGD-DDKTDIVFLDQVHIADNGTFTSLPT 276
Qy 285 FTFGTSATKDLNAQWAP--PLV-----NGLPQFSANFVAGIQLPLDFGI-----KSQON 333
Db 277 F-----ASKTLE-FDPSTKGLVYIPGGDIVQDTAN-----QKFYLEVDLTPESGPSEIN 324
Qy 334 MWAGAPASAAATGTDIGKLPMMPIQTSSG--NSTARN-----GSSSTRYSQDGPQ 384
Db 325 DPNDTSEYLNKLGAKUGSETNKKIYVGGILQNVNVIQNSYITQHAADFVMTDQDGYAR 384
Qy 385 GLVDVDTITSE-GKLGKQKYSNSQWDFYNIPLARFTSEDLGLRRGNHYSATLDSGDPF 443
Db 385 GELIDLVLVLEDGVDVGVVYSGNETLPTRYRLAQAQTDPEELVKKGSNLYASVKT---PTI 441

A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: flgE; BU340

Query Match 16.0%; Score 413.5; DB 2; Length 405;
Best Local Similarity 25.7%; Pred. No. 1.8e-18;
Matches 131; Conservative 73; Mismatches 185; Indels 121; Gaps 13;
QY 7 IGATGKTHSTGLGTGVSNNIANANTIGYKQOVVFDLFSQDLAIGSTGSGQGNQAGMA 66
DB 2 IASGLLANNDYMEIISNNIANASTIGYKSKPLFFDFMFSHSFVSNTN-----GYGV 54
QY 67 QVGSVRITFTOGAPEPGNSVTDLAIGKGFPO-VLEDKVHYTRAGNFRFTQDGLNDPS 125
DB 55 GISSIIIONFNGMLVETGRDLGIIKDGFPRLVDSQGHVYTRDQGFLLDKQNIINIQ 114
QY 126 GFTLMGSRIS-----NNPNIKKTEPIQLDFDNPVAKSPAKTSTALNAVNLG--- 175
DB 115 GMYLTGLNTSCSKSDFNN-----RNLFPINL--KSNILKXKPTSEIMLKAFLENRTESK 168
QY 176 ---DSTDKTQSEANPYFALLESWKNGTPTTSTSYAOPMRVYDQGNSHDITVYFDG 232
DB 169 SSVDSNDKLSKPEDYTIYS-----SFLTFNDGDELTSNVFNITSKDS 205
QY 233 APSSTGKTFEYLAMNPSEDSASGSDSAGLLMSGTMTFSSNGELKXMTAFTPTGSAT 292
DB 206 -KKTNKWTN--VESNDSDDKETKN-----SFLTFNDGDELTSNVFNITSKDS 254
QY 293 KDLNAWQAPLVLNGLPQFSANFVGAGIQPLTLDGFIKSGQQNMWAGAPASAAAIGTDIGKL 352
DB 255 KK-----YENITLNL----- 264
QY 353 PSMPIQTSSNGNSTARNSSSTRYSQDGYPOGLVDVITITSEKLOGKYSNQVDFYN 412
DB 265 -----TGTRISQSDVSWEHSQNGVPGQNLKTFDIVTNGEIIIGTYCNQKQQTIGQ 315
QY 413 IPLARFTSEDLRREGNNHYSATLDSGPEGL-PGTSNYKGLSVNOLETSNVDMSEMV 471
DB 316 ILLSKFINPEKLPESGNLWSATAESGEAKTAMKAGIQESGVLNKTLEVSNDVNLNELI 375
QY 472 NMIIIRGFQNMNSVTTADTLMQALELK 501
DB 376 NMIIAQRYQNSAQSFKTEKDIINTLINLQ 405

RESULT 18
DB2106
flagellar hook protein FlgE VC2197 [imported] - Vibrio cholerae (strain N16961 serogroup
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: DB2106
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, W.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: DB2106
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <HEI>
A;Cross-references: UNIPROT:Q9KQ10; GB:AE004292; NID:g9656760; PIDN:AAF9534
A;Experimental source: serogroup O1; strain N16961; biotype B1 Tor
C;Genetics:
A;Gene: VC2197
A;Map position: 1

Query Match 15.0%; Score 388; DB 2; Length 434;
Best Local Similarity 25.7%; Pred. No. 7.5e-17;
Matches 132; Conservative 69; Mismatches 213; Indels 100; Gaps 17;
QY 6 FIGATGKTHSTGLGTGVSNNIANANTIGYKQOVVFDLFSQDLAIGSTGSGQGNQAGMG 65

DB 3 YVLSGLSAAQMDLNTTNNIANANTFGKSRAEFGDVYSTSLF-----TNAKTTPQGG 57
QY 66 AQVGSVRITFTOGAPEPGNSVTDLAIGKGFPOVLEDKV-----HYTRAGNFRFTQDGL 121
DB 58 VQAAKVAQOFHEGSIYTNPLDLRIAGTGFFAVAKDRLVPOQNELTRNGAFHLDKNSFM 117
QY 122 NDPQSFTLMGSRISNNP---NIKKETLEPIQL--DFNPVAKSPAKTSTALNAVNLGD 176
DB 118 VTANDEFLLGYEV--NPDTGDVLSYEPKPINIPPOFGKP---KOTANIDLGANLSPAN-GD 171
QY 177 STDKTQSEANPYFALLESWKNGTPTTSTSYAOPMRVYDQGNSHDITVYFDGAPSS 236
DB 172 LKDP-----ALFDI-----TDP---ETYNRTSTSIYDSMGOPKYLTYYLKDYNQ 214
QY 237 TGSKTFEYLAMNPSE-----DGSAASTDSAGLLMSGTMTFSSNGELKXMTAFTPTGS 290
DB 215 ANTWQTYTVDTKTEKPIINVVGDAASPTGHVG---HTRFNNDGLTSS----- 261
QY 291 ATKDLNANQP---APLVNGLPQFSANFVGAGIQPLTLDGFIKSGQQNMWAGAPASAAAIGT 347
DB 262 ---LNGQPIVTEPL-----GGANPVDLNL-----GA 285
QY 348 DIGKLPSPMPIQTSSNGNSTARNSSSTRYSQDGYPOGLVDVITITSEKLOGKYSNQV 407
DB 286 DYNQTLSE-----SLDSATQPAAPFELTKFDQDGTATGTLTKIDFDENGSLVLTYSNGIN 340
QY 408 VDFYNIPLARFTSEDLRREGNNHYSATLDSGPEGLPGTSNYKGLSVNOLETSNVDM 467
DB 341 TTLGRVALVRVANEQGLDKKGTQMDATQFSKAKIWSGSKGSGFSGISNGSLEQSNIDMT 400
QY 468 REMANMIIIRGFQNMNSVTTADTLMQALELK 501
DB 401 QELVDLISQNRFOANSRALEVHNGLOQNILQIR 434

RESULT 19
EB1271
flagellar hook chain protein Cj1729c [imported] - Campylobacter jejuni (strain NCTC 11168
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: EB1271
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrall
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hysE
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: EB1271
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-865 <PAR>
A;Cross-references: UNIPROT:Q9PLU9; GB:AL139079; GB:AL111168; NID:g6968971; PIDN:CAB7371;
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: flgE2; Cj1729c

Query Match 14.9%; Score 384.5; DB 2; Length 865;
Best Local Similarity 19.9%; Pred. No. 3.4e-16;
Matches 172; Conservative 100; Mismatches 229; Indels 365; Gaps 23;
QY 1 MMGLFIGATGKTHSTGLGTGVSNNIANANTIGYKQOVVFDLFSQDLAIGST-----G 55
DB 1 MMRLSLMSGVSGLAQHADVGNLISNNTTGGFKYSRADFGTMTFSQTVKIATATPDGRG 60
QY 56 SQCPNQAQMGAVGVSVRTITFQCAPEPGNSVTDLAIGKGFPOVLEDKV--HYTRAGNF 113
DB 61 GSNFLQIGLGVSVSSTRIHSQGSVQTTDKNTDVAINGDGFPMVSDGGLTNYLTSRSGDF 120
QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKKTEPIQLDFDNPVAKSPAKTSTA-----L 168
DB 121 KLDAYGNFVNAGFVVQGNINWDDQITDSRTP-QNIFIDPGMHIPAAKSTEVAKANL 179
QY 169 NAVNLGDST-----DKTQ-----SEANPYFALL 192

Db 180 NSGLAIGTSRNLYALDSVHGWNKTQRAEDENDGTGTTQFYTTSKNSVEVTEKGVDAAGSL 239
Qy 193 ESWKNG-----
Db 240 FNAKQGLNLRDQGIWVSYADATYSTNKGVAFPDPLNQQNTAAFWGTANQKVNLDIT 299
Qy 200 -----TPPISTSNYSYAQPMRVYDOOG----- 221
Db 300 LNVRIQNRADIOSIDDAIAYINTFTPTDTRDGTGKAVKNGKSGIDFVNDNADGITDN 359
Qy 222 -----NSHDIITYPD-----GAP-----SSTGSK-----TF- 242
Db 360 MKNINLVANTTAGELMNAVNNNNQTFNNNGNGOAGTPTINKNGSSLWTATNITFT 419
Qy 243 -----EVLVAMN-----PSED 253
Db 420 POPQAATNVQLTGGLNAQIITAHKIIYSSNPVDIGPMNPDGPAFGPGANATRTPEP 479
Qy 254 GSAAS- GTDSAGLLMSGTWTFSSNGELKXN----- 282
Db 480 GSAAYWDAVNGGLLNTNVRTFTEDRELLQRDARYGVYDGGTFAAADINQIKVVV 539
Qy 283 -----TAFTPTGSATK-----DLNAWQ----- 299
Db 540 TADGHFAISNAEQSTVFPNAINGVGNATTTDPRKMSFNITAYSNKQGTSTNDATFAIF 599
Qy 300 ---PAPLVNG-----LPOFSANFVAGI-----QPLTLDFFGIKQQN 333
Db 600 KAFDGLVIGNQIKESQELKLSAFSAGLEIYDSLGSKHTLEVQFKQSTTQDGGNEWQMI 659
Qy 334 MWAGAPAGAAIG-----TDIGKLPMP-----IOTSSGN 364
Db 660 IRVPEPAEINTTGEPPNIIIVGTARFNDGSLASTPTPTNFPNNGAAPPNQIKLSFGT 719
Qy 365 S-----TARNGSSSTRYSODYPOGLVD--VITISEGKLOGKYSQSVVDFYNIPLA 416
Db 720 SGNNDGLVSSNSASLTQATDGTYSNGLKPDARVDDKGNILGEFTNGKTPFAVAKIAMA 779
Qy 417 RFTSDGLRRGNHYSATLDSGGPEFGLPGTSGNYKLSVNOLETNSVDMREMNMI 476
Db 780 SVANNNGLEEIGNLFKVTANSNIVVGEAGTGGRGEMKTSALEMSNVDLRSRLTELI 839
Qy 477 QRGFOWNSKSVTTADTLMQKALEKR 502
Db 840 QRGYQANSKRTISTDQMLQTLQLKQ 865

RESULT 20
B87361
flagellar hook protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: B87361
R.; Nierman, W.C.; DeBlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolof
n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87361
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-591 <STO>
A:Cross-references: UNIPROT:P35806; GB:AE005673; NID:gl3422168; PIDN:AAK22886.1; GSPDB:G
C:Genetics:
A:Superfamily: rod protein flag

Query Match 14.7%; Score 380.5; DB 2; Length 591;
Best Local Similarity 22.1%; Pred. No. 3.4e-16;
Matches 137; Conservative 98; Mismatches 230; Indels 155; Gaps 15;
Qy 4 SLFIGATGKMTHTSLGLTVSSNNIANANTIGYKQQVVFQDLFSQDLAIGTSGSGPNQAG 63

Db 6 AMLAGVSLIANSSALAAISDNIANVNTVGFKRSTSNFSTLV-----TSGNKQIOTYSA 58
Qy 64 MGAQVGSVRTITFTQGAPEFGNSVTDLAIGGKGF-----QVTLEDKVHYTRAGNERFT 116
Db 59 GGKQATHQFIQQGLTQSTTNLDISISGAGFFVTTEKPENLITATDTRSFTRAGSFQDL 118
Qy 117 QDGLNDPSGFTLMG-----SRISNNPIK 141
Db 119 NLGRLNDAGLVQGLADPVVSLITPDPSDLMLQLASINVSVGGTAETKTVGVNANLR 178
Qy 142 KETLEFIQDFNDPTVAKSPAKTISTALNAVNLG-DSTDKTQSEANPYFA---LLESWKG 197
Db 179 SEQFVAAAYSYKVT-AGSPSKTNVVDVSDATNHNHVDVYSSTGIANPVSGNNEYLVDIKE 237
Qy 198 NG---TPPI-----STSNYSYAQPM----- 214
Db 238 NGVIVATGVAVDAATNELVSTIDYKGPVTSMTTTRINAAGTTVNLADLGIYNASG 297
Qy 215 -----RVVD-----QQNSHDIITYFDGAPSTGSKTFFVLVAMNP-- 250
Db 298 ADDEAVVAGKLYDPSTWMSDYAKDMSKGVKDPDFEVOIPLSDSKGQRTVTLMLKGPQP 357
Qy 251 -----SEDSASAGTDSAGLLMSGTWTFSSNGELKXMTAFTPTGSAATKDLNWAQAP 302
Db 358 NOWYAELEKRAKPGDLAN--NNGQISTGIIIEFTIDGKLNKGTSLGFTTPT-----AI 407
Qy 303 LVNGLPQFSANFVAGIQPLTLDFFGIKQQNMMWAGAPASAAAIGTDIGKLPMPPIOTSS 362
Db 408 TIKSSGYIAPTTPPAVQPPT-----PPTWADALGIDQEV--QIDLASAA 451
Qy 363 GNSTARNGSSSTRYSODYPOGLVDVITISEGKLOGKYSQSVVDFYNIPLAFTSED 422
Db 452 GGLTQYNSQSVQSVNTNGTAFGLNLTNIEIDGGVSAIFDNGVTRRIAQVAIAFTSPN 511
Qy 423 GLRREGNHYSTLDSGGPEFGLPGTSGNYKLSVNOLETNSVDMREMNMI 482
Db 512 GLKGVNGNAYRVNTESSGYSLKAPSGGAGALAPSTLEASTVDLSQETGLITTOQAYSA 571
Qy 483 NKSQVTTADTLMQKALEKR 502
Db 572 SSKIITADQMLEELNKR 591

RESULT 21
A83956
flagellar hook protein flgE [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: A83956
R.; Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A83956
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-263 <STO>
A:Cross-references: UNIPROT:Q9KA41; GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BA0061
A:Experimental source: strain C-125
C:Genetics:
A:Gene: flgE
C:Superfamily: rod protein flgF

Query Match 14.3%; Score 370.5; DB 2; Length 263;
Best Local Similarity 22.6%; Pred. No. 4.5e-16;
Matches 115; Conservative 43; Mismatches 99; Indels 251; Gaps 9;
Qy 1 MWSLFIQATGKMTHTSLGLTVSSNNIANANTIGYKQQVVFQDLFSQDLAIGTSGT----- 54
Db 1 MLRSYSGISGMRNFTKLDVIGNNANVNTFGYKGRFTFDQMVVSQOLA-GATAPGANR 59
Qy 55 GSGQPNQAGMGAQVSVRTITFTQGAPEFGNSVTDLAIGGKGFQVFTLEDKVHYTRAGNFR 114


```
Db 60 GGVPKQVGLQGLASIDNVHVGSGQNTNRELDLIGSGDGFQVQDGTFFYTSRGNFY 119
QY 115 FTQDGLFNDPDSFTLMSGRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNL 174
Db 120 FDQGG-----TIVN- 128
QY 175 GDSTDKTQSEANPYFALLBSWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGAP 234
Db 129 -----AQGLRMDTNGN--PIVV----- 144
QY 235 SSTGSKTFEYLVAAMPSEDSAGTSDSAGLLMSGTMTFSSNGELKNTAFPTGSAKND 294
Db 145 -----PTAQSFSGID-----GTVV----- 161
QY 295 LNAWQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAGTIDICKLPS 354
Db 162 ----- 161
QY 355 MPMIQTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGKLGKYSNSQVVDYFNIP 414
Db 162 -----IDEDGDLQ-----EAGMYR 175
QY 415 LAETSEDGLRRGNHYSATLSDSGPEFGLPQTSNYGKLSVNQLETSNVDMSSREVMNMI 474
Db 176 LVNFANPEGLAKAGANLYLVANSAGAPLFTSAGNGGTGDIAGTLEMSNVVD;SEEFTEMI 235
QY 475 IIOGFGOMNSKSVTTADTLMQKALELKR 502
Db 236 VAQRGQFQANTRIITTSDEILQELVNLKR 263

RESULT 22
AI2646
Flagellar hook protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AI2646
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AI2646
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-425 <KUR>
A:Cross-references: UNIPROT:Q8UHV5; GB:AB008688; PIDN:AAL41591.1; PID:G17738927; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: flgF
A:Map position: circular chromosome

Query Match 14.2%; Score 367.5; DB 2; Length 425;
Best Local Similarity 26.2%; Pred. No. 1.4e-15;
Matches 135; Conservative 66; Mismatches 208; Indels 107; Gaps 13;

QY 1 MMGSLFIGATGMKTHSTGLGTVSNNIANANTIGYKQQQVWFQDLFSQDLAIGSTGSGPN 60
Db 3 IFGTWRTGVSGNNAQAANKLGIVGDNIANASTTGYKRASTSFSSL-----VLPSSSGS---- 54
QY 61 QAGMGAQVGSVRTITFQGAPEPGNSVTDLAIGGKGFQVLTLEDKVHY-TRAGNFRFTQDG 119
Db 55 YASGGVQSNVRSYISEQGNLSYTTSDLAIQNGGFFVVDGSGGTPYLITRAGSFQKNSEG 114
QY 120 FLNDPDSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTD 179
Db 115 YLENAAGQLMGYPYGSNP-----PA-----AVVN----- 139
QY 180 KTOSEANPYFALLEBSWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGAPSTGS 239
Db 140 -----GFTGLEA-----INVNFNFG-----LTASPSTQGS 163
QY 240 KTFEYLVAAMPSEDSAGTSDSAGLLMSGTMTFSSNGELKNTAF-----T 286
Db 164 -----FPANLRDNDKAATAP-----LPSGNSATAAFGNKTSILTAFDGSGAKVLYDFYTT 212
QY 287 PTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAG 346
Db 213 KTGDNTWEVAVYRQDQSTNGGFPYTAT-PAANLVQTKVDLEPDPATNKLTASPSTITID 271
QY 347 TDIKLPSPMPQIOTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGKLGKYSNSQ 406
```

```
Db 140 -----GFTGLEA-----INVNFNFG-----LTASPSTQGS 163
QY 240 KTFEYLVAAMPSEDSAGTSDSAGLLMSGTMTFSSNGELKNTAF-----T 286
Db 164 -----FPANLRDNDKAATAP-----LPSGNSATAAFGNKTSILTAFDGSGAKVLYDFYTT 212
QY 287 PTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAG 346
Db 213 KTGDNTWEVAVYRQDQSTNGGFPYTAT-PAANLVQTKVDLEPDPATNKLTASPSTITID 271
QY 347 TDIKLPSPMPQIOTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGKLGKYSNSQ 406

RESULT 23
H97428
FigE protein (AF325192) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: H97428
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Boughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H97428
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-425 <KUR>
A:Cross-references: UNIPROT:Q8UHV5; GB:AB007869; PIDN:AAK86385.1; PID:G15155515; GSPDB:G
C:Genetics:
A:Gene: AGR_C1012
A:Map position: circular chromosome

Query Match 14.2%; Score 367.5; DB 2; Length 425;
Best Local Similarity 26.2%; Pred. No. 1.4e-15;
Matches 135; Conservative 66; Mismatches 208; Indels 107; Gaps 13;

QY 1 MMGSLFIGATGMKTHSTGLGTVSNNIANANTIGYKQQQVWFQDLFSQDLAIGSTGSGPN 60
Db 3 IFGTWRTGVSGNNAQAANKLGIVGDNIANASTTGYKRASTSFSSL-----VLPSSSGS---- 54
QY 61 QAGMGAQVGSVRTITFQGAPEPGNSVTDLAIGGKGFQVLTLEDKVHY-TRAGNFRFTQDG 119
Db 55 YASGGVQSNVRSYISEQGNLSYTTSDLAIQNGGFFVVDGSGGTPYLITRAGSFQKNSEG 114
QY 120 FLNDPDSGFTLMGSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAVNLGSDTD 179
Db 115 YLENAAGQLMGYPYGSNP-----PA-----AVVN----- 139
QY 180 KTOSEANPYFALLEBSWKNGTTPPISTNSYSAQPMRVYDQGNSHDITVYFDGAPSTGS 239
Db 140 -----GFTGLEA-----INVNFNFG-----LTASPSTQGS 163
QY 240 KTFEYLVAAMPSEDSAGTSDSAGLLMSGTMTFSSNGELKNTAF-----T 286
Db 164 -----FPANLRDNDKAATAP-----LPSGNSATAAFGNKTSILTAFDGSGAKVLYDFYTT 212
QY 287 PTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNWAGAPASAAAAG 346
Db 213 KTGDNTWEVAVYRQDQSTNGGFPYTAT-PAANLVQTKVDLEPDPATNKLTASPSTITID 271
QY 347 TDIKLPSPMPQIOTSSGNSSTARNSSSTRYSQDGYPOGDLVDVITITSEGKLGKYSNSQ 406
```

Db 272 DQVSGVQAINIDLSQMTQFSKFTPGFAVINGPSQ--TKDVEIGKDGGLVTVAYQDGG 329
QY 407 VVDYFNIPLARFTSDGLRENNHYSATLSGGGPEFGLPGTSGNYGKLSVNNQLETSNDVM 466
Db 330 RNNIYQALATVPVSDNLTQNGVYLPNSDGVVITIGFQSGSGFYQKGALEGSNDVI 389
QY 467 SREWNVNIIRGFQFOMNSKSVTTADTMQKALELKR 502
Db 390 ASELTDIESQRIYANSKVFQSGDLMDVLINLKR 425
RESULT 24
D64633
Flagellar hook - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C/Accession: D64633
R/Tomb, J.F.; White, O.; Kexlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: D64633
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-605 <rom>
A:Cross-references: UNIPROT:Q25566; GB:AE000600; GB:AE000511; NID:Q2314042; PIDN:AAD0795

Query Match 13.7%; Score 354.5; DB 2; Length 605;
Best Local Similarity 23.5%; Pred. No. 1.5e-14;
Matches 151; Conservative 82; Mismatches 222; Indels 177; Gaps 22;
QY 1 MMGSLFIGATGKTHSTGLTVSNINANTIGYKQQQVFDLFSQDLAIGTSGGPN 60
Db 1 MNDTLNAYSIGIKTHQFGIDSLSNINANTILGYRNDPEKTLFSSHLDAKSVAN 60
QY 61 QAGGAQVGSVRTIPTQ-GAPEPGNSVTDLAIGKGF--OVTL-----EDK 104
Db 61 DRNYGV-TGSGNVLSNKDGEYMPGEFHMAYQKGFVIGPNKNGEITINKDGFSSKQD 119
QY 105 VHYTAGNFRFTQDGLNDPSGFTLMG-----SRISNNPNKK--ETLPIQL 150
Db 120 NFLTRAGNFARDADGYLVTPEGVYVYDGLKKIKDGLNSTARDEDIEKLHGLNLSPLQI 179
QY 151 -----DF-----NDP 155
Db 180 PQDLTYQPVLSKTVNISVNLNPKDHLKGVQDFLNDKGEIKERFLNQDINALANDNEP 239
QY 156 TVAKSPAKTSTAL-----NAVNLGD-----STDKTQSEANPYFA 190
Db 240 IDAITNRKLNISIQEKDGKEDFVFTYGDAGKGNQFKTLGDLQKLLKERTGLDLN--- 295
QY 191 LLESWKNGTPTPI-----STSNYSVAQPMRVYDQOGSHD-ITVYFDGAPSTGSKTFEYVAMN 249
Db 296 LIKSEKDAKSPALLEIANPSETPIAFSLSGGIADKGLANQWELKKGISRDS--VAIK 353
QY 250 PSEDCSAAAGTDSAG---LLMSGTMTFSSNGELKNMTAFTPTGSAT--KDLNAWQAPAPLV 304
Db 354 IPYVSTEVDIYDKAGDKYLLQSEYMTNSN-----DPTSPSTSKRKNQWTWVKSXI 404
QY 305 -----NGLPQPSANFVGAGIQ-----PLTLDGFIKSOQNMWAGAPASAAIGTIDIGK 351
Db 405 ADPNKKTPIINDPTWEIVGFDSATHKMSAPMTLDF-----KGNK 443
QY 352 LPSMMPIQTSNGSTARNSSSTRYSDGVPQDGLVDVITITSEGKLOGKYSNSQVVDY 411
Db 444 LTVSLDKSENHSDLSYQDSKLEASQDGKPRGIFEDMRIBENGIVISAFSGVVEPVA 503
QY 412 NIPLARTSDEGLRENNHY---SATLD-----SCGPPEGL--PGTNSYKLSVNNQLE 460

Db 504 RIGILAFNTDQGLRKIGGNLYEMQSGTNGENRPLSGNPILGWDEBEGKLKFGKIRHKYLE 563
QY 461 TSNVDMRSRENVNMIIRGFQFOMNSKSVTTADTMQKALELKR 502
Db 564 TSNVNAGNALTNLILMQRGYSNNARAFGAGDDMIKEAISLKK 605
RESULT 25
G71881
probable flagellar basal-body/rod/hookprotein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C/Accession: G71881
R/Alm, R.A.; Ling, D.S.L.; Moir, D.T.; King, B.L.; Brown, B.D.; Doig, P.C.; Smith, D.R.;
ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; J
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: G71881
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-605 <arn>
A:Cross-references: UNIPROT:Q9ZKU1; GB:AE001513; GB:AE001439; NID:94155407; PIDN:AAD06422
A:Experimental source: strain J99
A:Genetics:
A:Gene: jhp0844

Query Match 13.5%; Score 349.5; DB 2; Length 605;
Best Local Similarity 22.8%; Pred. No. 3e-14;
Matches 148; Conservative 86; Mismatches 224; Indels 191; Gaps 21;
QY 1 MMGSLFIGATGKTHSTGLTVSNINANTIGYKQQQVFDLFSQDLAIGTSGGPN 60
Db 1 MNDTLNAYSIGIKTHQFGIDSLSNINANTILGYRNDPEKTLFSSHLDAKSVAN 60
QY 61 QAGGAQVGSVRTIPTQ-GAPEPGNSVTDLAIGKGFQV-----TLEDK 104
Db 61 DRNYGV-TGSGNVLSNKDGEYMPGEFHMAYQKGFVIGPNKNGEITINKDGFSSKQD 119
QY 105 VHYTAGNFRFTQDGLNDPSGFTLMG-----SRISNNPNKK--ETLPIQL 150
Db 120 NFLTRAGNFARDADGYLVTPEGVYVYDGLKKIKDGLNSTARDEDIEKLHGLNLSPLQI 179
QY 151 -----DF-----NDP 155
Db 180 PQDLTYQPVLSKTVNISVNLNPKDHLKGVQDFLNDKGEIKERFLNQDINALANDNEP 239
QY 156 TVAKSPAKTSTAL-----NAVNLGD-----STDKTQSEANPYFA 190
Db 240 IDAITNRKLNISIQEKDGKEDFVFTYGDAGKGNQFKTLGDLQKLLKERTGLDLN--- 295
QY 191 LLESWKNGTPTPI-----STSNYSVAQPMRVYDQOG--NSHDITVYFDGAPSTGSKTF 242
Db 296 LIKSEKDAKSPALLEIANPSETPIAFSLSGGIADKGLANQWELKKGISRDSVAIKIP 355
QY 243 EYLVAMPSEDCSAAAGTDSAG---LLMSGTMTFSSNGELKNMTAFTPTGSAT--KDLNA 297
Db 356 YY-----STEVDIYDKAGDKYLLQSEYMTNSN-----DPTSPSTSKRKNQ 397
QY 298 WOPAPLV---NGLPQPSANFVGAGIQ-----PLTLDGFIKSOQNMWAGAPASAAA 344
Db 398 WEVKSXIYDPPNKTPIINDPTWEIVGFDSATHKMSAPMTLDF----- 439
QY 345 IGTIDIGKLPMMPIQTSNGSTARNSSSTRYSDGVPQDGLVDVITITSEGKLOGKYSN 404
Db 440 ---KGNKLTYSLDKSENHSDLSYQDSKLEASQDGKPRGIFEDMRIBENGIVISAFSN 496
QY 405 SQVDFYNIPLARTSDEGLRENNHY---SATLD-----SCGPPEGL--PGTNSYK 453
Db 497 GVVEPVARIGLAFINDOGLRKIGNLYEMQSGTNGENRPLSGNPILGWDEBEGKLKFGK 556
QY 454 LSVNQLETSNVDMRENVNMIIRGFQFOMNSKSVTTADTMQKALELKR 502

A:Reference number: A42365; MUID:91258343; PMID:1828465
A:Accession: I42365
A:Molecule type: DNA
A:Residues: 1-149,'S',151-264 <ALB>
A:Accession: S14502
A:Molecule type: DNA
A:Residues: 1-264 <AL2>
A:Cross-references: EMBL:X56049; NID:G39904; PIDN:CAA39528.1; PID:G39912
E:Zuberi, A.R.; Ying, C.; Bischoff, D.S.; Ordal, G.W.
Gene 101, 23-31, 1991
A:Title: Gene-protein relationships in the flagellar hook-basal body complex of Bacillus
A:Reference number: JG0019; MUID:91285431; PMID:1905667
A:Accession: JG0023
A:Molecule type: DNA
A:Residues: 1-149,'S',151-264 <ZUB>
A:Cross-references: GB:M54966; NID:g142918; PIDN:AAA22444.1; PID:g142919
C:Genetics:
A:Gene: flgB; flgG
C:Superfamily: rod protein flgP

Query Match 12.9%; Score 333; DB 2; Length 264;
Best Local Similarity 22.0%; Pred. No. 9.8e-14;
Matches 112; Conservative 43; Mismatches 103; Indels 250; Gaps 9;

Qy 1 MMGLFICATGMKTHSTGLGTVSSNNIANANTIGYKQQVVFOQLFSQDLAIGST-----G 55
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 MLRSLYSGISGMKNFOTKLVDIGNNIANVTVGFKSRVTFKDMVSQTLAGGSAAAGATIG 60
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 56 SQGFNOAGMGQAQGVSVRIFTOGAPCGNSVDTLAITGGKGFFOVTTLEDKVHYTRAGNFRF 115
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 61 GTNSKQIILGSSSGTIDITHSTASQTOSTGRTLDAIDGDGYFRIDTGCDGTAYTRAGNFYL 120
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 116 TDQGLNDPSGFTLMGSRISNNPNIKKETLEPTQLDFNDFTVAKSPAKTSTALNAVYNLG 175
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 121 DNTGLVLTGDGY-----HVLNMN 138
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 176 DSTDKTQSSEANPYFALLESWKNGTPPISTSNYSYAQPWRVYDQGNSHDITYFDGAPS 235
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 139 GGTIKIPTDA----- 148
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 236 STGSKTFEYLVAMNPSEDGSAASGTDSDAGILMSGTWTFSSNGELKNMFTAFTPTGSATKOL 295
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 149 -----QTFS----- 152
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 296 NAWQPAPLVNLGPQSFANFVGAGIOPLTLDFGIKSQQNMWAGAPASAAATGTDIGKLPSM 355
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 153 -----IGSD-GKV--- 159
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db	149	-----QTFS-----	152
Qy	296	NAWQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQNNWAGAPAGAAAIGTDIGKLPMS	355
Db	153	-----IGSD-GKV-----	159

QY 356 MFIQTSSNGNSTARNSSSTRYSQDYPQGLVDVITITSEGLKQKYSNSQVDFYNIPL 415
Db 160 -----SIVD-----AEGTKQ-----DGGQIGI 176
QY 416 ABFTSEDGLRREGNNHYSATLDG-GPEFGLPGTSNYGKLSVNQLETNSVDMSEMNMI 474
Db 177 VTFANSDGLDKGSLNLYRESLSGTFASAAOPDGGTGALKSGFLENSVNDLTDEFTMI 236
QY 475 IIQRGFQMSKSVTTADTLMQKALELKR 502
Db 237 VAQRGFQMSKIITTSDEIQELVNLKR 264
RESULT 28
B72243
flagellar basal-body rod protein FlgG - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: B72243
R.Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: B72243
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-261 <ARN>
A:Cross-references: UNIPROT:Q9X1M8; GB:AE001800; NID:g4982090; PIDN:AAD366Q
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1542
C:Superfamily: rod protein flgF
Query Match 12.6%; Score 325.5; DB 2; Length 261;
Best Local Similarity 22.4%; Pred. No. 2.8e-13;
Matches 114; Conservative 37; Mismatches 103; Indels 255; Gaps 8;
QY 1 MMSLFIGATGKTHSTGLGTVSNNIANANTIGYKQOVVFDLFSQDLAIGSTGSGPN 60
Db 1 MMSLISATGMSAQCFKLDITIANLANVDVTGYKRAEFDLLYQYVKNAGTPTAATS 60
QY 61 QAGMGAQVG-SVRT-----IFTQGAFFPGNSVTDLAIGGKGFQVTLDE-KVHYTRAGNF 113
Db 61 SLPTGLYVGHGVRAATRTIFLNGFQGTGALDLAGDGFQIQDGRYATRDGSP 120
QY 114 RFTQDGFNDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAV 173
Db 121 KMDSEG-----RIVTSN----- 126
QY 174 LGDSTDKTQSEANPYFALLESWKNGTPTISTSNYSVAQPMRVYDQGNSHDITVYFDGA 233
Db 127 -----RIVTSN----- 132
QY 234 PSSTGSKTFEYLVAMPNSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTTGSATK 293
Db 133 -----GLLIVPEITI----- 142
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTDFGIKQQNMWAGAPASAAAIGTDIGKLP 353
Db 143 -----PENAVSINV----- 151
QY 354 SMPIQTSSNGNSTARNSSSTRYSQDYPQGLVDVITITSEGLKQKYSNSQVDFYNI 413
Db 152 -----SPDGIVSALQDGTIQELG-----TI 172
QY 414 PLARFTSEDGLRREGNNHYSATLDGSGPERGLPCTSNYKLSVNQLETNSVDMSEMNMI 473
Db 173 TLVRFVNPGLKISGIDNLYIATASGDPTEGVPGQDGFAGIKQGFLEKSNVDVVRMVDN 232
QY 474 IIQRGFQMSKSVTTADTLMQKALELKR 502

Db 233 ITAQRAYEFNSRVIQTADENMLRTATNKR 261
RESULT 29
C70372
flagellar hook basal body protein flgG - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: C70372
R.Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: C70372
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <AQF>
A:Cross-references: UNIPROT:Q67006; GB:AE000709; NID:g2983373; PIDN:AAC06963.1; PID:g2983
A:Experimental source: strain VFS
C:Genetics:
A:Gene: flgG2
C:Function:
A:Description: located in the distal rod
C:Superfamily: rod protein flgF
C:Keywords: basal body; flagellum
Query Match 12.0%; Score 311; DB 2; Length 262;
Best Local Similarity 21.2%; Pred. No. 2.3e-12;
Matches 108; Conservative 45; Mismatches 102; Indels 254; Gaps 9;
QY 1 MMSLFIGATGKTHSTGLGTVSNNIANANTIGYKQOVVFDLFSQDL-----AIGSTGS 56
Db 1 MPRALWTSAGMTAQTDLNLSHMANVNTVYKMKMATFODLLYQYVEPSPGAPTPT 60
QY 57 QGFN--QAGMGAQVGSVRTIFTQGAFFPGNSVTDLAIGGKGFQVTLDE-KVHYTRAGNF 113
Db 61 RSPSGFQIGLGYVSDTYGIFTQGNLTKTNDQLDIAIQDGGFFKVLDPDGTIAVTRNGQF 120
QY 114 RFTQDGFNDPSGFTLMGSRISNNPNIKETLEPIQLDFNDPTVAKSPAKTSTALNAV 173
Db 121 RLDSGRIVNDSGYL-----DP----- 138
QY 174 LGDSTDKTQSEANPYFALLESWKNGTPTISTSNYSVAQPMRVYDQGNSHDITVYFDGA 233
Db 139 -----EITL----- 142
QY 234 PSSTGSKTFEYLVAMPNSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTTGSATK 293
Db 143 ----- 142
QY 294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTDFGIKQQNMWAGAPASAAAIGTDIGKLP 353
Db 143 -----PADAISV----- 149
QY 354 SMPIQTSSNGNSTARNSSSTRYSQDYPQGLVDVITITSEGLKQKYSNSQVDFYNI 413
Db 150 -----SIAADGTVSVLR-----QGA--TEVEEVGRI 173
QY 414 PLARFTSEDGLRREGNNHYSATLDGSGPERGLPCTSNYKLSVNQLETNSVDMSEMNMI 473
Db 174 ELAKFVNPAGLRRIGNNLYIQTATASGPTIDNPGQGLTLLQGLYLESSNNVIVEMVNL 233
QY 474 IIQRGFQMSKSVTTADTLMQKALELKR 502
Db 234 IIAQRAVEFNTKGIQAADMLSQAAANLRR 262
RESULT 30
HB3510
flagellar basal-body rod protein FlgG PA1082 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C;Accession: H83510
R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83510
A;Molecule type: DNA
A;Status: preliminary
A;Residues: 1-261 <STO>
A;Cross-references: UNIPROT:Q914P7; GB:AE004539; GB:AE004091; NID:g9946985; PIDN:AAG0447
A;Experimental source: strain PA01
C;Genetics:
C;Superfamily: rod protein flgF

Query Match 11.9%; Score 307; DB 2; Length 261;
Best Local Similarity 21.6%; Pred. No. 4e-12;
Matches 108; Conservative 46; Mismatches 93; Indels 254; Gaps 11;

QY 1 MMGSLFICATGKTHSTGLGTVSNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGPN 60
DB 1 MSLALWVSKTGLSAQDNVLTISNLANVSTTGKDRABEFQDLVYQIRPQGGQSTQDS 60

QY 61 QAGMGAQVGS-VRTIFQGAPEPGNSVT-----DLAIGKGFQFQVLEL-KVHYTRAGNF 113
DB 61 ELPSGLQGTGVRVVGKTFPGSLQTTQPLDMVANGGFQVLLPDGTVSYTRDGSF 120

QY 114 RFTQDGLNDPSGFTLMGSRISNNPNIKKTELEPIQLDNDPTVAKSPAKTSTALNAV 173
DB 121 HLNSDQGVITSGFAL-----EPAL----- 140

QY 174 LGDSTDKTQSEANPYFALLESKWKGNGTPISTSNYSYAQPMRVYDQGNSHDITVYFDGA 233
DB 141 ----- 140

QY 234 PSSTGSKTFEYLVAMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSATK 293
DB 141 -----VVENE-----TOTET----- 150

QY 294 DLNAQAPLVNGLPQSFANFVGAGIQPLTLDPGIKSQQNNWAGAPASAAAIGTDIGKLP 353
DB 151 -----VG----- 152

QY 354 SMPIQTSSGNSSTARNSSSTRYSQDGPQGLVDVITITSEGLQKYSNSQVDFYNI 413
DB 153 -----QDG-----TVSVTTGNAQ-----PQVIG--NI 173

QY 414 PLARFTSDGLRRGNHNSATLDSGGPEFGLPQTSNYGKLSVNQLTNSVNDMSRMVNM 473
DB 174 QTADFIPNAGLQAGNNLFLETGSSGAPQVGTGPGINGLGTVAQNTLENSNVNVEELVNM 233

QY 474 IILQRFQFQNSKSVTTADTML 494
DB 234 ITTORAYEMNSKVIADQML 254

RESULT 31
H80810
flagellar basal body rod protein flgG [imported] - Escherichia coli (strain O157:H7, sub
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: H90810
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A93629; MUID:21156231; PMID:11258796
A;Accession: H90810
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <HAY>

A;Cross-references: UNIPROT:P75939; GB:BA000007; PIDN:BA834879.1; PID:gl3360920; GSPDB:G
A;Experimental source: strain O157:H7, substrain RMD 0509952
C;Genetics:
A;Gene: ECSI456
C;Superfamily: rod protein flgF

Query Match 11.7%; Score 303; DB 2; Length 260;
Best Local Similarity 21.4%; Pred. No. 7e-12;
Matches 109; Conservative 44; Mismatches 98; Indels 258; Gaps 10;

QY 1 MMGSLFICATGKTHSTGLGTVSNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGPN 60
DB 1 MISSLIAKTGLDAAQQTMDVIANNLANVSTNGFQRQAVFEDLLYQ--TIROPQAQSE 58

QY 61 QAGM--GAQVGS-VETIFT-----QGAPEPGNSVTDLAIGKGFQFQVLEL-KVHYTRAG 111
DB 59 QTTLPGLQGTGVRVPAERLHSGQNSLQTNNSKDVAKGQFQVWLPDGSSTAYTRDG 118

QY 112 NFRFTQDGLNDPSGFTLMGSRISNNPNIKKTELEPIQLDNDPTVAKSPAKTSTALNAV 171
DB 119 SFQVDQNGQLVTAGGFQV-----QPAITIPA----- 144

QY 172 VNLGSDTKTQSEANPYFALLESKWKGNGTPISTSNYSYAQPMRVYDQGNSHDITVYFD 231
DB 145 ----- 154

QY 232 GAPSTGSKTFEYLVAMNPSEDGSAAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSA 291
DB 155 GVSVYV-----QCGQAA----- 166

QY 292 TKDLNAQAPLVNGLPQSFANFVGAGIQPLTLDPGIKSQQNNWAGAPASAAAIGTDIGK 351
DB 167 ----- 166

QY 352 LPSMMPLOTSSGNSSTARNSSSTRYSQDGPQGLVDVITITSEGLQKYSNSQVDFY 411
DB 167 -----PVQVQQLN----- 174

QY 412 NPLARFTSDGLRRGNHNSATLDSGGPEFGLPQTSNYGKLSVNQLTNSVNDMSRMV 471
DB 175 ---LTFMNDTGLSEIGENLYTETQSSGAPNESTFGLNGAGLLYQGYVETSNVNAEELV 231

QY 472 NMILIQRFQFQNSKSVTTADTMLQKALEL 500
DB 232 NMIVQRAYEINSKAVSTTDQMLQKLTQL 260

RESULT 32
D85670
flagellar basal body rod protein flgG [similarity] - Escherichia coli (strain O157:H7, s
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: D85670
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potancousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85670
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <STO>
A;Cross-references: UNIPROT:P75939; GB:AE005174; NID:gl2514619; PIDN:AAG55824.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: flgG
C;Superfamily: rod protein flgF

Query Match 11.7%; Score 303; DB 2; Length 260;
Best Local Similarity 21.4%; Pred. No. 7e-12;
Matches 109; Conservative 44; Mismatches 98; Indels 258; Gaps 10;

QY 1 MMGSLFICATGKTHSTGLGTVSNNIANANTIGYKQQVVFQDLFSQDLAIGSTGSGPN 60

```
Db 1 MISSLWIAKTGLDAQCTNMVIANLANVSTNGFRQRAVFEDLLYQ--TIROPGAQSS 58
QY 61 QAGM--GAQVGS-VRTIFT-----QCAPEPGNSVTDLAIGKGFQVLTLED-KVHYTRAG 111
Db 59 QTTLPGLQIGTGVPRVATERLHSGNLSQTNNSKDVAIKGGFFQVMLPDGSSAYTRDG 118
QY 112 NFRFTQDGLNDPDSGFTLMSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAV 171
Db 119 SFQVDQNGQLVTAGGFQV-----OPAITIPA----- 144
QY 172 VNLGSDTKTQSEANPYFALLSWKNGTPTISTNSYAOQMRVYDQGNSHDITVYFD 231
Db 145 -----NALSITIGRD 154
QY 232 GAPSTGSKTFEYLVAIMPSEDSASAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSA 291
Db 155 GVSVT-----QQQQA----- 166
QY 292 TKDLNAWQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNMWAGAPASAAAIGTDIGK 351
Db 167 ----- 174
QY 352 LPSMMPLOTSSNGSTARNSSSTRYSQDGPQGLDVTITSEKLGKQKYSNSQVDFY 411
Db 167 -----PVQVQQLN----- 174
QY 412 NIPLARFTSEGLRREGNNHYSATLDSGPEFGLPGTSNYGKLSVNQLETSNVDMDSREMY 471
Db 175 ---LTTFMNDTGLSEIGENLYTETQSSGAPNESTFGLNGAGLLYQGYVETSNVNVAEELV 231
QY 472 NMIIQRFQMSKSVTTADTMLQKALEL 500
Db 232 NMIOVRAVEINSKAVSTTDQMLQKLTQL 260
RESULT 33
C64851
flagellar basal body rod protein flgg - Escherichia coli (strain K-12)
N:Alternate names: distal rod protein
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64851
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-260 <BLAT>
A:Cross-references: UNIPROT:P75939; GB:AE000208; GB:U00096; NID:g1787308; PIDN:AAC74162.
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: flgG; flaL; fla_VVII
C:Function:
A:Description: located in the distal rod
C:Superfamily: rod protein flgg
C:Keywords: basal body; flagellum
Query Match 11.7%; Score 303; DB 2; Length 260;
Best Local Similarity 21.4%; Pred. No. 7e-12;
Matches 109; Conservative 44; Mismatches 98; Indels 258; Gaps 10;
QY 1 MWGSLFIGATGKTHSTGLGTVSNNIANANTIGYKQOVFQDLFSQDLAIGSTGSGPN 60
Db 1 MISSLWIAKTGLDAQCTNMVIANLANVSTNGFRQRAVFEDLLYQ--TIROPGAQSS 58
QY 61 QAGM--GAQVGS-VRTIFT-----QCAPEPGNSVTDLAIGKGFQVLTLED-KVHYTRAG 111
Db 59 QTTLPGLQIGTGVPRVATERLHSGNLSQTNNSKDVAIKGGFFQVMLPDGSSAYTRDG 118
QY 112 NFRFTQDGLNDPDSGFTLMSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAV 171
Db 119 SFQVDQNGQLVTAGGFQV-----OPAITIPA----- 144
QY 172 VNLGSDTKTQSEANPYFALLSWKNGTPTISTNSYAOQMRVYDQGNSHDITVYFD 231
Db 145 -----NALSITIGRD 154
QY 232 GAPSTGSKTFEYLVAIMPSEDSASAGTDSAGLLMSGTMTFSSNGELKNMTAFTPTGSA 291
Db 155 GVSVT-----QQQQA----- 166
QY 292 TKDLNAWQAPLVNGLPQFSANFVGAGIQLTLDFGIKSQQNMWAGAPASAAAIGTDIGK 351
Db 167 ----- 174
QY 352 LPSMMPLOTSSNGSTARNSSSTRYSQDGPQGLDVTITSEKLGKQKYSNSQVDFY 411
Db 167 -----PVQVQQLN----- 174
QY 412 NIPLARFTSEGLRREGNNHYSATLDSGPEFGLPGTSNYGKLSVNQLETSNVDMDSREMY 471
Db 175 ---LTTFMNDTGLSEIGENLYTETQSSGAPNESTFGLNGAGLLYQGYVETSNVNVAEELV 231
QY 472 NMIIQRFQMSKSVTTADTMLQKALEL 500
Db 232 NMIOVRAVEINSKAVSTTDQMLQKLTQL 260
RESULT 34
XMBEFG
flagellar basal body rod protein flgg - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: S08174; S15356
R:Omura, M.; Kutsukake, K.; Hasebe, M.; Iino, T.; Macnab, R.M.
J. Mol. Biol. 211, 465-477, 1990
A:Title: FlgB, FlgC, FlgF and FlgG. A family of structurally related proteins in the flag
A:Reference number: S08171; MUID:90172414; PMID:2129540
A:Accession: S08174
A:Molecule type: DNA
A:Residues: 1-260 <HOM>
A:Cross-references: UNIPROT:P16439; EMBL:X52094; NID:g47674; PIDN:CAA36314.1; PID:g47677
R:Jones, C.J.; Macnab, R.M.; Okino, H.; Aizawa, S.I.
J. Mol. Biol. 212, 377-387, 1990
A:Title: Stoichiometric analysis of the flagellar hook-(basal-body) complex of Salmonella
A:Reference number: S15353; MUID:90204563; PMID:2181149
A:Accession: S15356
A:Molecule type: protein
A:Residues: 1-5 <JON>
C:Genetics:
A:Gene: flgG
C:Function:
A:Description: located in the distal rod
C:Superfamily: rod protein flgg
C:Keywords: basal body; flagellum
Query Match 11.6%; Score 301; DB 1; Length 260;
Best Local Similarity 21.4%; Pred. No. 9.4e-12;
Matches 109; Conservative 43; Mismatches 99; Indels 258; Gaps 10;
QY 1 MWGSLFIGATGKTHSTGLGTVSNNIANANTIGYKQOVFQDLFSQDLAIGSTGSGPN 60
Db 1 MISSLWIAKTGLDAQCTNMVIANLANVSTNGFRQRAVFEDLLYQ--TIROPGAQSS 58
QY 61 QAGM--GAQVGS-VRTIFT-----QCAPEPGNSVTDLAIGKGFQVLTLED-KVHYTRAG 111
Db 59 QTTLPGLQIGTGVPRVATERLHSGNLSQTNNSKDVAIKGGFFQVMLPDGSSAYTRDG 118
QY 112 NFRFTQDGLNDPDSGFTLMSRISNNPNIKKETLEPIQLDFNDPTVAKSPAKTSTALNAV 171
Db 119 SFQVDQNGQLVTAGGFQV-----OPAITIPA----- 144
```

QY 172 VNLGDSGDTQSEANFYFALLESWKNGTTPISTSNYSYAQPMRVYDQOQNGSHDITVYFD 231
Db 145 -----NALSITIGRD 154
QY 232 GAPSSGSKTFEYLVAMNPSDGSAAAGTDSAGLLMSGTWTSSNGELKNTAFTPTGSA 291
Db 155 GVVSVT-----QQGQAA-----166
QY 292 TKDLNAWQAPLVNGLPQFSANFVGAIQPLTDLFGIKSQNNWAGAPASAAAIGTDIGK 351
Db 167 -----166
QY 352 LPSMFIQTSSNGSTARNSSSTRYSQDGYPOGDLVDVTITSEGLQKGYNSQVVDYF 411
Db 167 -----PVQVQQLN-----174
QY 412 NIPLARFTSEGLRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETSNVDMREMY 471
Db 175 ---LTFPMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGYVETSNNVABELV 231
QY 472 NMIIQRFQFOMNSKSVTTADTMLQKALEL 500
Db 167 -----166
QY 352 LPSMFIQTSSNGSTARNSSSTRYSQDGYPOGDLVDVTITSEGLQKGYNSQVVDYF 411
Db 167 -----PVQVQQLN-----174
QY 412 NIPLARFTSEGLRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETSNVDMREMY 471
Db 175 ---LTFPMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGYVETSNNVABELV 231
QY 472 NMIIQRFQFOMNSKSVTTADTMLQKALEL 500
Db 232 NMIIQVQRAYEINSAVSTTDQMLQKLTQL 260
RESULT 35
AD0640
flagellar basal-body rod protein FlgG (distal rod protein) [imported] - Salmonella enter
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AD0640
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AD0640
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08303.1; PID:G16502349; GSPDB:GN00176
C:Genetics:
A:Gene: STY1218
C:Superfamily: rod protein flgF
Query Match 11.6%; Score 301; DB 2; Length 260;
Best Local Similarity 21.4%; Pred. No. 9.4e-12;
Matches 109; Conservative 43; Mismatches 99; Indels 258; Gaps 10;
QY 1 MMGSLFIGATGKTHSTGLGTVSNNIANANTIGYKQQQVFDLPSQDLAIGTSGQGN 60
Db 1 MISLLWIAKTGLDAQOTNMVIANLANVSTNGFKQRAVFEDLLYQ--TIROFGAQSSE 58
QY 61 QAGM--GAQVGS--VRTTFT-----QGAPEPNSVTDLAIGGKGFQVFTLEDKVH--YTRAG 111
Db 59 QTTIPSGLIQIGTVRPAERLHQQNLSTNNKDVAKGQGFQVWLPDGTSAVTRDG 118
QY 112 NFRFTQDGLNDPSPGFTLMGSRISNNPNIKETLEP:QLDNDPTVAKSPAKSTALNAV 171
Db 119 SFQVDQNGQLVTAGGFQV-----QPAITIPA-----144
QY 172 VNLGDSGDTQSEANFYFALLESWKNGTTPISTSNYSYAQPMRVYDQOQNGSHDITVYFD 231
Db 145 -----NALSITIGRD 154
QY 232 GAPSSGSKTFEYLVAMNPSDGSAAAGTDSAGLLMSGTWTSSNGELKNTAFTPTGSA 291
Db 155 GVVSVT-----QQGQAA-----166
QY 292 TKDLNAWQAPLVNGLPQFSANFVGAIQPLTDLFGIKSQNNWAGAPASAAAIGTDIGK 351

Db 167 -----166
QY 352 LPSMFIQTSSNGSTARNSSSTRYSQDGYPOGDLVDVTITSEGLQKGYNSQVVDYF 411
Db 167 -----PVQVQQLN-----174
QY 412 NIPLARFTSEGLRRGNHYSATLDSGGPEFGLPGTSNYGKLSVNQLETSNVDMREMY 471
Db 175 ---LTFPMNDTGLESIGENLYIETQSSGAPNESTPGLNGAGLLYQGYVETSNNVABELV 231
QY 472 NMIIQRFQFOMNSKSVTTADTMLQKALEL 500
Db 232 NMIIQVQRAYEINSAVSTTDQMLQKLTQL 260
RESULT 36
E97165
flagellar hook protein FlgE. [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: E97165
R:Noelling, J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A36900; MUID:21359325; PMID:21359325
A:Accession: E97165
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-423 <KUR>
A:Cross-references: UNIPROT:Q97H59; GB:AE001437; PIDN:AAK80112.1; PID:G15025147; GSPDB:GN
A:Experimental source: Clostridium acetobutylicum ATCC924
C:Genetics:
A:Gene: CAC2154
Query Match 11.3%; Score 291.5; DB 2; Length 423;
Best Local Similarity 20.0%; Pred. No. 7.4e-11;
Matches 111; Conservative 73; Mismatches 187; Indels 183; Gaps 14;
QY 1 MMGSLFIGATGKTHSTGLGTVSNNIANANTIGYKQQQVFDLPSQDL--AIGSTGSG 58
Db 1 MLRANTSGISGLKVNQNKLDIVGNSIANSSTAYKQSINFSDAISETIKGASGATANYG 60
QY 59 ---PNOAGMGAQVGSVRTFTQGAPEPNSVTDLAI--GKGGFQV-----TLEDK-----104
Db 61 GYNPEQITLVNSSAIITNTASGVNTTGRNLDVAITDGDGVFMVASCANMADESECIN 120
QY 105 -----VHTRAGNFRPTQDGLNDPSPGFTLMGSRISNNPNIKETLEPIQ 149
Db 121 STTDSHTSTPSGTDISTYRDSGNLDNEGNLVTADGRKVLGYSM-----VGRNTL-----171
QY 150 LDFNDPTVAKSPAKSTALNAVNLGDSGDTQSEANFYFALLESWKNGTTPISTSNYS 209
Db 172 ---YED-----SADTKNYIYGNFVSIQADET-----196
QY 210 YAPQMRVYDQOQNGSHDITVYFDGAPSSGSKTFEYLVAMNPSDGSAAAGTDSAGLLMSG 269
Db 197 -----TMAGSDENAAATYIAAEGGDDLLNKLSIAAIVKDPNPSGSGTATIKKGDVAFVDAN 248
QY 270 TMTFSSNGELKNTAFTPT---GSATKDLNAWQAPLVNGLPQFSANFVGAIQPLTLD 326
Db 249 DVDLRADNRYLHTLKLIPPTVKVTATRD-----276
QY 327 GIKSQNNWAGAPASAAAIGTDIGKLPSPMPITQSSGNSTARNSSSTRYSQDGYPOD 386
Db 277 -----ASGNV-----YSTQELKVKGF-----293
QY 387 LVDVTITSEGLQKGYNSQVVDVFNIPLARFTSEGLRRGNHYSATLDSGGPEFGLP 446
Db 294 ----SIGQDGVITANLDNGSAAAIQGIAMSFNTAGGLEKGGNYKPSASSGQPMWRSG 349
QY 447 GTSN-----YGLKLSVNQLETSNVDMREMYNNMIIIQRFQFOMNSKSVT 488

Db 350 AYSNKLDPTFISSAKKEDNSGGYQWYSGALEMSNVDLAQQFSDMIVASRAYQANGKIIT 409
QY 489 TADTWMQKALELKR 502
Db 410 TSDEILQDLVNLKR 423

RESULT 37
AC2644
Flagellar basal-body rod protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AC2644
R:Wood, D.W.; Satubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, M.; E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:21608550; PMID:11743193
A:Accession: AC2644
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <KUR>
A:Cross-references: UNIPROT:Q44338; GB:AE008688; PIDN:AAL41569.1; PID:g17738903; GSPDB:G17738903
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: flgG
A:Map position: circular chromosome
A:Superfamily: rod protein flgG

Query Match 11.2%; Score 290.5; DB 2; Length 262;
Best Local Similarity 21.5%; Pred. No. 4.3e-11;
Matches 109; Conservative 44; Mismatches 86; Indels 267; Gaps 12;

Db 2 MGSLEFGATGKTHSTGLTGVSNNTANANTIGYQQQVVDLPFSDLAIGSTGSGQ-PN 60
1 MRALAAATGMDAQQTLEVIANNIANTTGYKRAAEFTDLYQ-----TERMGGVFN 55

QY 61 QA-----GMGAQVGSVRTIFTQG-AFEPGNSVTDLAIGGKGFQVTLSD-KVHY 107
56 RANQAIPEGANIGLVQTSVAVRNIHQNLITGKGL-DVAIIQGQWFQIEAADGSTLY 114

QY 108 TRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKKETTLEPIQLDFNDPTVAKSPAKTSTA 167
115 SRAGAFNKNADGNLVTVDGYNVI-----PNI-----NIPT----- 144

QY 168 LNAVNLGDSDTKQSEANPYFALLESWKNGTPTISTNSYSAQPMRVYDQGNSHDIT 227
145 -----TRGQVTAIGNAAD--FTQLG----- 150

QY 228 VYFDGAPSGTSGKTFEYLVAWNPSEDSAGSTDSAGLLMSGTMTFSSNGELKNMTAFTP 287
151 I----- 151

QY 288 TGSATKDLNAWQAPLNVNGLPQPSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGT 347
152 ----- 151

QY 348 DIGKLPNMPQTSSGNSSTARRYSQDGYPGQDLVDVTITSEGLKQKYSNSQV 407
152 -----TRGQVTAIGNAAD--FTQLG----- 171

QY 408 VDFYNIPLARTSDGRLRRGNHYSATLDSGGPEFGLPCTSNYKLSVNQLETNSVDMS 467
151 I----- 151

QY 468 REMYNMIIQRFQNMKSQVTTADTM 493
228 KEITDLITAGRAYEMNSKVITTADEM 253

RESULT 39
F71801
Flagellar basal-body rod protein (distal rod protein) - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: F71801
R:Alm, R.A.; Ling, R.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999

RESULT 38
C97426
Flagellar basal-body rod protein flgG (distal rod protein) [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: C97426
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, D.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:21608551; PMID:11743194
A:Accession: C97426
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <KUR>
A:Cross-references: UNIPROT:Q44338; GB:AE007869; PIDN:AAK86364.1; PID:g15155490; GSPDB:G15155490
C:Genetics:
A:Gene: AGR_C_972
A:Map position: circular chromosome
C:Superfamily: rod protein flgG

Query Match 11.2%; Score 290.5; DB 2; Length 262;
Best Local Similarity 21.5%; Pred. No. 4.3e-11;
Matches 109; Conservative 44; Mismatches 86; Indels 267; Gaps 12;

QY 2 MGSLEFGATGKTHSTGLTGVSNNTANANTIGYQQQVVDLPFSDLAIGSTGSGQ-PN 60
1 MRALAAATGMDAQQTLEVIANNIANTTGYKRAAEFTDLYQ-----TERMGGVFN 55

QY 61 QA-----GMGAQVGSVRTIFTQG-AFEPGNSVTDLAIGGKGFQVTLSD-KVHY 107
56 RANQAIPEGANIGLVQTSVAVRNIHQNLITGKGL-DVAIIQGQWFQIEAADGSTLY 114

QY 108 TRAGNFRFTQDGLNDPSGFTLMGSRISNNPNIKKETTLEPIQLDFNDPTVAKSPAKTSTA 167
115 SRAGAFNKNADGNLVTVDGYNVI-----PNI-----NIPT----- 144

QY 168 LNAVNLGDSDTKQSEANPYFALLESWKNGTPTISTNSYSAQPMRVYDQGNSHDIT 227
145 -----TRGQVTAIGNAAD--FTQLG----- 150

QY 228 VYFDGAPSGTSGKTFEYLVAWNPSEDSAGSTDSAGLLMSGTMTFSSNGELKNMTAFTP 287
151 I----- 151

QY 288 TGSATKDLNAWQAPLNVNGLPQPSANFVGAGIQPLTLDGFIKSOQNMWAGAPASAAAIGT 347
152 ----- 151

QY 348 DIGKLPNMPQTSSGNSSTARRYSQDGYPGQDLVDVTITSEGLKQKYSNSQV 407
152 -----TRGQVTAIGNAAD--FTQLG----- 171

QY 408 VDFYNIPLARTSDGRLRRGNHYSATLDSGGPEFGLPCTSNYKLSVNQLETNSVDMS 467
151 I----- 151

QY 468 REMYNMIIQRFQNMKSQVTTADTM 493
228 KEITDLITAGRAYEMNSKVITTADEM 253

RESULT 39
F71801
Flagellar basal-body rod protein (distal rod protein) - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: F71801
R:Alm, R.A.; Ling, R.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999

A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: F71801
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <ARN>
A;Cross-references: UNIPROT:Q9ZJ26; GB:AE001570; GB:AE001439; NID:94156108; PIDN:AAD0706
A;Experimental source: strain J99
C;Genetics:
A;Gene: flgG.2
C;Superfamily: rod protein flgP

Query Match 11.1%; Score 288; DB 2; Length 262;
Best Local Similarity 21.0%; Pred. No. 6.1e-11;
Matches 107; Conservative 39; Mismatches 109; Indels 254; Gaps 9;
1 MMSLFICATGKTHSTGLGTGVSNNIANANTIGYKQQVVFQDLFSDLAIGSTGSGQGN 60
1 MLRSLSYATSGMLAQTHIITNNIANVNTTGFKSRADFNDLFYQACVAGTNTSNT 60
61 -----QAGGAQGVSVRTITQGAPEPGNSVTDLAIGKGFQVLTED-KVHYTRAGNF 113
61 LSPDGMVGLGVAPSAITKMFQSGPKETENNLDIAITGKGFQVLPDGTATYTRSGNF 120
114 RFTQDGFNDPFGFTLMGSRISNNPNIKKTELEPIQLDFNDPTVAKSPAKTSTALNAVYN 173
121 KLDQGLVTSQGVLLI-----PQI---TL-----PEDTQO-----VN 150
174 LQDSTDKTQSEANPYFALLESWKNGTPPISTNSYAPQPMRYVYDQGNSHDITVYFDGA 233
151 IG-----VDGT 156
234 PSSTGSKTFYLVAMNPSEDSASGTDTSAGLLMSGTMTFSSNGELKNWTAFTPTGSATK 293
157 VSVT----- 160
294 DLNAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKISQQNWAGAPASAAAIGTDIGKLP 353
161 ----- 160
354 SMPDQSSNSTARNSSSTRYSQDGYPCQGLVDVITSEKLGKYSNSQVDFYNI 413
161 -----QGLQTTSNVIG--QI 173
414 PLAFPTSEDLRRBNHYSATLDSGGPEGLPQTSNYGKLSVNQLSTSNVDMSREVMNM 473
174 TLANFVNPAGLHSGNDLFAITNASGDVAVGNPDQSLGKLRQGFLELSNVLVEENTDL 233
474 IIIQRFQMSKSVTTTADTMLQKALELKR 502
234 ITAQRAYEANSKSTQADAMLQTVNSLKR 262

Query Match 11.1%; Score 287.5; DB 2; Length 454;
Best Local Similarity 25.7%; Pred. No. 1.5e-10;
Matches 124; Conservative 59; Mismatches 169; Indels 131; Gaps 19;
1 MMSLFICATGKTHSTGLGTGVSNNIANANTIGYKQQVVFQDLFSDLAIGSTGSGQGN 54
1 MLRSLSYATSGMLAQTHIITNNIANVNTTGFKSRADFNDLFYQACVAGTNTSNT 60
55 GSQGNQAGMGAQGVSVRTITQGAPEPGNSVTDLAIGKGFQVLTED-KVHYTRAGNF 112
61 GQNDVSVGLGVNNTTKVFSQSTQNTDVKTLAIEGCGFFIISPRGKTQFTRDGE 120
113 RFTQDGFNDPFGFTLMG-----SRISNNPNIKKTELEPIQLDFNDPTVAKSPAK 163
121 PLFDANGNLVTNGGVVQGMKDDLRNAEKISEDDFRVDTTKPIQNIQIDPAM-MMPAR 179
164 TST--ALNAVNLGSDTKTQSEANPYFALLESWK--NGTPPISTNSYAPQPMR----- 215
180 ASSNISLEANLNAGRHVDQV---AN-VFGLSGTTPVDGINPIYDSDHNLTKAEDFGA 235
216 VYDQGN-----SHDITVYFDGAPSGTSGKTFEYLVMNPSEDSASGTDTS- 262
236 LFTQSGDAIGLTENOGIWWVSYKTSWVNDIEATSGESSIEINHTRISPTNDSASGSISSV 295
263 -----AGLLMSGTMTFSSNG-----ELKN----- 281
296 VAAQNAINALKQKTGVEAFVNDGMLRLQNKNNMDGAEVKNIRITADGTGAFANFIEGDS 355
282 -MTAF---TPTGSATKDLNAWQAPLVNGLPQFSANFVGAGIQPLTLDGFIKISQQNWMA 336
356 DITAFRYRTTISASPDSTGQGFRTEDLRALIQYDANLIDKPSQAYT----- 402
337 GAPASAAAIGTDIGKLPSPMPPIQT-----SSGNSTARNSSSTRYS 378
403 ---DSTASVVKFNKY-GMLEIQNKNDGDELKQDLNIFVSGYSSENST-NNGS-----E 451
379 QDG 381
452 QDG 454

Search completed: October 26, 2004, 09:10:16
Job time : 48 secs

RESULT 40
S61455
flagellar hook protein flgE - Helicobacter mustelae (fragment)
C;Species: Helicobacter mustelae
C;Date: 19-Mar-1997 #sequence_revision 17-Jul-1998 #text_change 18-Sep-1998
C;Accession: S61455; S61444
R;O'Toole, P.W.; Kosrzynska, M.; Trust, T.J.
Mol. Microbiol. 14, 691-703, 1994
A;Title: Non-motile mutants of Helicobacter pylori and Helicobacter mustelae defective
A;Reference number: S61442; MUID:95198543; PMID:7891557
A;Accession: S61455
A;Molecule type: protein
A;Residues: 1-41;420-454 <OTO>
A;Experimental source: strain 4298
A;Accession: S61444
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 31-39,'S',41-420 <OWO>
A;Cross-references: EMBL:U09548
A;Experimental source: strain 4298

